

Pursuing arrogant simplicities

Multidisciplinary research in biology requires the patience to distinguish untutored crassness from deceptively simple insights, and awareness from all participants of just how complex is even the simplest life-form.

Commenting on the impact of physicists on biology after the Second World War, the physicist-turned-biologist Leo Szilard said: "What physicists brought to biology is not any skills acquired in physics, but rather an attitude: the conviction that few biologists had at that time, that mysteries can be solved." Physicists always tend to start simple. Is that wise, when confronting life?

Let no one underestimate the irritation that hypothetical simplicity can engender. Just consider the reactions to 'Daisyworld' (A. J. Watson & J. E. Lovelock, *Tellus* **35B**, 284–289; 1983). Jim Lovelock, the father of Gaia theory, described this model as "my proudest invention", but it continues to annoy biologists and Earth scientists to this day.

Lovelock's fictional planet covered with black and white daisies, adjusting its reflectivity and thus its temperature under a changing sun, was his answer to critics of his proposal that the combination of a planet and its life-forms could be self-regulating. The irritation that the model has engendered stems from both a proper scientific scepticism about its relevance to Earth, but also, at times, an improper defensiveness of traditional disciplinary boundaries.

In its own complex way, post-genomic biology may be misleadingly simplistic about regulation. Generating vast sets of data from stressed cells in order to determine patterns of gene expression is an immense step forward. But beware the false impression that we are close to understanding how networks of genes regulate one another's expression, and generate phenotypes such as cellular development and behaviour.

Even the true scale of most genetic networks is unknown. And biologists know that genes are just one aspect of control: protein switches and molecular signalling networks are still a largely uncatalogued universe. It could be well over a decade before we might start to predict how self-contained subsets of networks — 'modules' — within a cell will behave under a specified perturbation, although qualities such

as the modules' robustness are beginning to yield to network analysis.

How can the skills of physicists, chemists and engineers be brought to bear on these problems? Telling them to 'get real' by reading a biology textbook is fair advice, but not adequate. Even after one absorbs a thousand or more pages of text, one would still be unlikely to have a feel for the variability and complexity of even the simplest microbe. (Whether many biologists have a genuine feel for the organisms they study is a pertinent question, come to that.)

Seemingly every top-notch university in the United States has somewhere on its campus a building, newly completed or under construction, that is to bring together scientists from many disciplines to address some aspect of biology (see page 256 for a profile of Harvard's contribution to the genre). To their credit, some, like Stanford's Bio-X, are placing a high priority on multidisciplinary education for undergraduate and graduate students, rather than simply pulling researchers together and hoping that sparks will fly.

But while non-biologists need to be appropriately humble in the face of life's realities, they also need the confident faith — arrogance, if you will — that such systems can be modelled and made predictable. But what if, as some biologists suggest, there may be no possible model simpler than life itself? Such are the defeatist speculations that physicists at least (as Szilard suggested) are schooled to ignore.

Especially in a multidisciplinary project, it's important to be able to distinguish between ignorant simple-mindedness and the simplicity of true insight. This is easily said but less easily done, as researchers, journal editors and funding agencies repeatedly discover. But those scientists of any discipline who immerse themselves in life's innards could yet find that deceptively simple hypotheses earn them respect from biologists who are drowning in data — even if that respect takes years of stubborn persistence to be earned. ■

In search of sound science

Quality standards for US federal research could be useful — but only if industry acts in good faith.

'Sound science' is one of those phrases repeated so often by politicians in recent years that it has come close to losing all meaning. In endless debates over this or that piece of environmental or health regulation, each side invokes 'sound science' to mean anything that supports its end of the argument.

Now US government agencies are being asked to bring more precision to the term. They will soon be required by law to ensure that the scientific information they disseminate meets standards for "quality, objectivity, utility, and integrity" (see page 249).

The move seems to have started as a manoeuvre by industry-backed groups — which helped to draft the law — to make it harder to enact federal regulations. But it could nevertheless prompt a useful discussion of what constitutes 'quality' research, and what the role of science should be in shaping policy.

In public discussion of guidelines for implementing the new law,

scientists have argued, perhaps defensively, that the current system of peer review assures quality control. But peer review isn't infallible (see page 258), and President George W. Bush's White House, which casts a wary eye on federal regulation, calls for an additional standard of 'reproducibility'. By this it means that data and research methods supporting 'influential' government actions should be transparent enough for others to achieve the same result using the same tools.

In a world where everyone acts in good faith to practice 'sound science', that standard might be reasonable. But if industry uses the new law simply to gum up the regulatory process with legal challenges, or keeps sending scientists running back to the lab for one more test before accepting their results, this would be far from reasonable. If that happens, scientists and the wider public should rightly demand another kind of transparency — one in which industry lobbyists are required to come clean about their motives. ■





Intelligent debate
Scientists hit out
at creationism in
schools
p250



Dioxin damage
US and Vietnam to
cooperate over legacy
of Agent Orange
p252



Election promise
French socialist
candidate sets sights
on science
p253



Dead ringer
Joke name for
dinosaur fails to
raise a smile
p254

Concern mounts as US agencies face challenges to data quality

Tony Reichhardt, Washington

US government agencies are scrambling to craft policies that will guarantee the quality of the data they produce, in compliance with a little-noticed law that was passed over a year ago.

The White House Office of Management and Budget (OMB) has given government agencies until 1 May to come up with guidelines for how they will implement the law, which many observers say will make it easier for industry and other affected groups to challenge federal regulations on scientific grounds.

Even though the new law (Public Law 106-554) does not come into force until October, the industry-backed Center for Regulatory Effectiveness (CRE) has already invoked it. The centre asked the White House science office to stop disseminating the National Climate Assessment, which synthesizes scientific thinking about the consequences of climate change, on the grounds that it relies on "inaccurate" computer models.

The CRE, whose advisory board includes several former OMB officials, helped to draft the data-quality provision of the law, a



Climate change may be taking its toll in Alaska, but a law could suppress an assessment of its impact.

mammoth spending bill covering many agencies that moved through Congress just before Christmas 2000.

The law directs agencies to find procedures for "ensuring and maximizing the quality, objectivity, utility, and integrity of information" that they disseminate, and to come up with mechanisms allowing "affected

persons to seek and obtain correction" of such information where necessary.

A wide range of information published by the government may be affected by the law, including lists of endangered species and research that supports the regulation of pollutants, such as arsenic or airborne soot. All will now have to meet new standards set

Grants for youth aim to revive 'dying' Russian science

Bryon MacWilliams, Moscow

Russian president Vladimir Putin has responded to mounting complaints about the dire state of the country's science by offering a new grant scheme to young scientists and their supervisors.

In an announcement on 13 March, Putin said that 600 scientists under 35 years of age will receive supplementary grants of 24,000 rubles (US\$770), with their supervisors receiving the same amount. The recipients will be selected through competition.

Putin hopes that the grants will help to slow the exodus of young scientists from Russia, and of young people from Russian science. The average scientist in Russia is now 52 years old, according to government statistics.

Although many researchers have derided the size of the grants — which are nonetheless worth as much as a typical Russian researcher's salary — others have welcomed the fact that Putin has at least acknowledged the problem.

Russian politicians are slowly turning their attention to the collapse of science in the country. For example, Sergei Mironov, speaker of the upper house of the Russian parliament, is backing a budget proposal that would more than double federal funding for science, from less than 2% of the country's total budget to about 4%.

"If the government doesn't focus attention on the fundamental sciences, they will soon die out," Mironov told an Earth sciences meeting in St Petersburg on 5 March.

Boris Kagarlitsky, a sociologist at the Institute of Comparative Politics in the Russian Academy of Sciences, warns that the new grants could undermine research still further, if they are awarded in lieu of a long-term strategy of investment in research.

And some scientists commenting on Lenta.ru, a discussion website, have derided the grants as insufficient. "The government does not have any money, and should find the strength to admit it," one anonymous correspondent said. "To pompously propose such small grants is hypocritical." However, others said that the grants are better than nothing, and would be helpful to researchers such as mathematicians, theoretical physicists and economists, who do not need expensive equipment.

by each agency as to what constitutes "quality" research.

In the year since the law was passed, the OMB has published draft guidelines that question whether peer review is, in itself, sufficient to guarantee quality. The OMB originally argued that it was not, citing "cases where flawed science has been published in respected journals".

In some cases, the OMB said, results must be "substantially reproducible" to pass muster. That raised eyebrows in some quarters, such as the National Academy of Sciences, whose president Bruce Alberts wrote in a public comment that reproducibility sets a "new and unreasonable standard" for disseminating scientific information.

The OMB backed off, but not entirely: "It is not OMB's intent that each agency must reproduce each analytic result before it is disseminated," the agency said in its final guidelines last month. But the data and methods used should be transparent enough for someone else to reproduce them.

Researchers fear that the law will be used unscrupulously by opponents of regulation to challenge findings, in disciplines such as climate or ecology, that may be difficult to replicate independently, as in cases where the opportunity to collect fresh data has passed.

Each agency is responsible for coming up with its own procedures for guaranteeing data quality and handling complaints. Jim Tozzi, a Washington-based regulatory consultant and CRE adviser, who signed the letter requesting the withdrawal of the National Climate Assessment, thinks the burden of proof will be on challengers.

"You have to have real data" to make a challenge, Tozzi says. But he predicts that if petitioners are not satisfied by an agency's response, disputes will end up in the courts.

Most federal agencies are still feeling their way to understanding the full implications of the law, according to Tozzi and others. The Environmental Protection Agency declined to comment on how it would draft its policies, and the White House science office, which received the CRE's complaint about the National Climate Assessment, says it is referring the matter to White House lawyers.

But at the Environmental Protection Agency and elsewhere, the level of concern is rising as the May deadline nears. The agency is inviting public comments online on the subject this week, when the National Academy of Sciences will also hold the first of a series of workshops on the law's implications.

♦ www.whitehouse.gov/omb/fedreg/reproducible.html

Evolution critics seek role for unseen hand in education



Creative thought: advocates and critics of 'intelligent design' cross swords in Columbus, Ohio.

Trisha Gura, Cleveland

The great debate on teaching evolution and creationism in American schools is back — this time with an added twist.

The Ohio State Board of Education is being pressed to instruct teachers in the mid-western state to include 'intelligent design' in their biology lessons, as a possible alternative to Darwin's theory of evolution.

Proponents of 'intelligent design' concede that evolution takes place, but argue that its outcome is too complex to have occurred by chance and so must have been designed by some unseen hand. They told a rowdy public hearing of the board in Columbus on 11 March that including the topic in the school curriculum would help students to understand that scientific theories are always open to challenge.

But many scientists regard 'intelligent design' as pseudoscience, and say that it is being used as a Trojan Horse to introduce the teaching of creationism into schools.

On the other side of the Atlantic, meanwhile, scientists were protesting because of reports that Emmanuel College in Gateshead — a Christian-run technical college near Newcastle upon Tyne — is indoctrinating its students with creationist ideas. Richard Dawkins, an evolutionary biologist at the University of Oxford, accused the college of teaching "ludicrous falsehoods" and called on the government's education department to investigate further.

In the United States, school curricula are controlled at the state level, and Ohio's education board is revising its standards in response to instructions from Bob Taft, its Republican governor. But at least three members of the 19-person board have taken exception to a draft of the standards, produced in consultation with scientists, because it fails to acknowledge 'intelligent design' as a rival theory to that of evolution.

In response to these concerns, a subcommittee of the board invited two advocates of 'intelligent design' — Stephen Meyer and Jonathan Wells, both fellows of the Discovery Institute, a conservative think tank based in Seattle, Washington — to debate the topic with two of its critics, theoretical physicist Lawrence Krauss of Case Western Reserve University in Cleveland and cell biologist Kenneth Miller of Brown University in Providence, Rhode Island.

Around 1,500 people attended the sometimes-heated debate, in which Wells and Meyer characterized Darwin's theory as being under fire from within the scientific community. Wells waved a list of what he described as 40 peer-reviewed papers criticizing darwinism, and Meyer asked the board to "just permit teachers to teach the evolution controversy".

Krauss said that there was no such controversy. "'Intelligent design' is an idea," he said. "It is not science, because it does not appear in any peer-reviewed literature." Miller and Krauss both dismissed 'intelligent design' as "creationism dressed up as science".

A subcommittee of the board is due to present standards for science, and other subjects under review, to the full board by September. The board is expected to implement the standards in December — just after elections for six of the board members.

Some observers of Ohio politics say that — with the tacit backing of Taft — the state might implement standards that will open the door to the teaching of 'intelligent design'. Biologists worry that such a move could force the topic into biology textbooks — and open the way for change in Texas, which will discuss its standards next year. "Ohio is just a skirmish," says Miller, a co-author of five biology textbooks. "But it is a rehearsal for what will happen later."

Missed US nomination leaves climate post up for grabs

Jim Giles, Washington

The future leadership of the Intergovernmental Panel on Climate Change is up in the air this week, after the United States failed to renominate Robert Watson, the panel's current chair.

Watson's six-year term as chair of the panel, which was set up by the United Nations in 1988 to establish an international consensus on climate-change research, ends this year. The panel's Geneva-based secretariat had asked member nations to put forward their nominations for the position by 15 March.

India's nomination of Rajendra Pachauri, director of the Tata Energy Research Institute in New Delhi and current vice-chair of the panel, was the only one to arrive by the deadline, secretariat officials say. Pachauri is a former adviser to the World Bank and a director of the Indian Oil Corporation.

Watson, an ecologist who is chief scientist at the World Bank, says that he would be willing to serve a second term as chair and that he still hopes the United States will support his re-election. Watson was nominated by the United States when he was elected in 1996, but observers say that his former role as associate director for environment at

the White House Office of Science and Technology Policy under President Bill Clinton, and his closeness to environmental advocacy groups, may dissuade the Bush administration from nominating him this time round.

Renate Christ, deputy secretary of the secretariat, says that there is no clear procedure for how nominations should be made, and that the secretariat may continue to accept nominations up until the election of the chair in April. She also points out that Bert Bolin, the Swedish meteorologist who preceded Watson as chair, was re-elected for his second term without being renominated.

The panel's governing body will vote on the position when it meets in Geneva in April. It is unclear how member states would vote if both Watson and Pachauri stand. Christ says that several countries have expressed a desire for Watson to stay on. But some nations may feel that it is time for a representative of a developing nation to head the panel.

Whoever is elected will have to oversee the panel's fourth climate-change assessment, which is due to be completed over the next five to seven years.

■ www.ipcc.ch

Breast-cancer survey sets screening age for women

Helen Pearson, London

The World Health Organization (WHO) has added its voice to the heated debate over when women should start having mammograms.

The International Agency for Research on Cancer (IARC) — part of WHO — convened a panel of 24 scientists this month to review the available evidence for the health benefits of mammography. Chaired by Bruce Armstrong of the University of Sydney, the panel concluded that women aged 50 and over benefit from breast-cancer screening using mammograms.

But it fell short of endorsing the recommendation made last month by the US health department that women over 40 should have regular mammograms (see *Nature* 415, 950; 2002).

Some researchers still question whether the total number of lives saved by screening outweigh those lost as a result of treatment — even in older women.

But the IARC hopes that its final report, to be published within three months, will help to inform policy-makers around the world who are implementing or refining screening programmes.

Cash shortfall means time out for physicists

Geoff Brumfiel, Washington

Nuclear physicists in the United States have a synchronization problem. Just as much-wanted operating time is becoming available on major facilities, the physics division of the National Science Foundation (NSF) is running out of cash to support the researchers who want to use the equipment.

This year, the funds available to the physics division for supporting primary investigator grants have fallen by 11.5% compared with 2001. Given that most of the money is already committed to grants lasting several years, this translates into a cut of one-third in the number of new grants to be awarded in the next couple of months.

Nuclear physicists claim that they will be hit especially hard, because they will miss a rare opportunity to run their best facilities at close to full capacity. Earlier this year, the Department of Energy proudly announced that the United States' two largest facilities for nuclear science — the Relativistic Heavy Ion Collider at Brookhaven National Laboratory in New York state and the Continuous Electron Beam Accelerator

Facility at the Thomas Jefferson National Accelerator Facility in Newport News, Virginia — will have their operating times substantially increased next year.

The NSF's physics division will receive a 4% budget increase this year. But because of its obligations to run its own facilities in Michigan, Louisiana and Washington state, and its increased grants to larger physics groups, the division is facing a sharp cut in the number of individual researchers it can support. "It's a difficult situation," concedes Joseph Dehmer, the division's director.

"It's going to have a very, very unfortunate effect," says Lawrence Cardman, associate director of physics at the Jefferson lab. Cardman says that over a quarter of the lab's staff are funded by the NSF, and many will lose their grants this year. He fears that young scientists will be forced to abandon nuclear physics as a result.

"This is a disastrous situation," says Baha Balantekin of the University of Wisconsin in Madison. Most graduate students are supported by primary investigator grants, he says, and nuclear physics is already running



Physicists may lack the funds needed to use Brookhaven's Relativistic Heavy Ion Collider.

short of them. "This is cutting off blood flow to the field," he says.

But Dehmer claims that the situation may improve next year. President George Bush's budget proposal allows for a 5% increase in NSF funding in 2003 — including a 1.3% reduction at the physics division — but the House of Representatives budget committee last week suggested a more generous 9% increase for the agency.

US and Vietnam join forces to count cost of Agent Orange

RICHARD VOGEL/AP



Pollution by defoliating chemicals has been blamed for causing birth defects among Vietnam's children.

David Cyranoski, Tokyo

A joint research programme to investigate the damage caused by the defoliant Agent Orange is set to be launched by the United States and Vietnam.

But with no special funding — and with some government officials on both sides said to be keen to keep details of the damage under wraps — some experts doubt whether the project can get to the bottom of the health and environmental problems caused by the chemical. Used by the US military in the Vietnam War, Agent Orange has left a legacy of dioxin pollution — which has been linked to cancer, immune disorders and birth defects — in some parts of Vietnam.

On 10 March, following a four-day conference in Hanoi which brought together epidemiologists, toxicologists and environmental scientists from 13 countries (see *Nature* **413**, 442; 2001), the United States and Vietnam signed a memo formally establishing the joint research programme.

Chris Portier, director of the environmental toxicology programme at the US National Institute of Environmental Health Sciences, says that he is encouraged by the agreement. "We have always had a good relationship with the Vietnamese scientists, and now the Vietnamese government has become more comfortable with us," he says.

Vietnamese researchers were likewise optimistic. "The scientists of the two countries will work together and will see the damages suffered by the Vietnamese people," says Hoang Trong Quynh, director of Vietnam's Center for Ecologically Sustainable Agriculture. The findings of the research will "strongly support a humanitarian reso-

lution from the US side", Quynh adds.

The agreement sets up a US–Vietnam advisory committee that will arrange researcher visits, meetings and exchanges of graduate students. But the lack of special funding for the programme means that researchers will have to apply for support through the usual channels at agencies such as the US National Institutes of Health.

Critics claim that the Vietnamese government placed restrictions on who could participate at the Hanoi meeting. According to one delegate, the government is worried that the monitoring of dioxin levels in fish and seafood caught off the coast of Vietnam might damage exports. As a result, this participant claims that "there were no open discussions" at the meeting.

This and other complaints about the joint effort are set to come under scrutiny at a meeting in Stockholm this July, which will be attended by many of those who went to Hanoi. Al Burke, editor of the Nordic News Network, who is organizing the Stockholm meeting, claims that "many legal and ethical issues are being avoided because the United States doesn't want to deal with them".

But researchers at the Hanoi meeting say that the joint programme will lead to better research on Agent Orange's health and environmental effects. Presentations at the conference examined the link between exposure to dioxins and problems such as birth defects, lymphoma, breast cancer, Hodgkin's disease and diabetes. And the US delegation returned from Hanoi with about 20 Vietnamese scientific articles which will be translated into English, and might form the basis for future research projects. ■

Diplomats near pact in simmering debate over transgenic foods

Jim Giles, Washington

The European Union (EU) and the United States have taken a small step back from a possible trade war over genetically modified foods.

In negotiations in Yokohama, Japan, which ended on 8 March, representatives from the two sides reached a compromise on how such foods should be monitored. They have just over a year left to reach agreement on other issues, such as how the foods should be labelled for consumers.

The negotiations were part of a three-year plan to update the Codex Alimentarius, the food-safety standards used by the World Trade Organization (WTO) to settle trade disputes. In Yokohama, the United States dropped its earlier resistance to a system to monitor the movement and health risks of transgenic crops.

The EU currently bans the farming of about 30 US transgenic crop varieties and the importation of foodstuffs containing them. Their suppliers want the United States to file a complaint about the ban with the WTO — but diplomats on both sides hope that such action can be averted by updating the Codex rules.

Alan Randall, secretary of the Rome-based Codex commission, warns that the argument is far from over. The two sides are still divided, for instance, over how to deal with foods that are obtained from transgenic plants but which do not themselves contain transferred genes. The EU wants consumers to be told if the food is from a genetically modified source.

Negotiations on such disagreements are scheduled to be completed before a Rome meeting of the Codex commission in July 2003. Randall says there is a long road ahead, but is pleased that an agreement has been reached. "At least the two parties are no longer just shouting at each other," he says. ■

♦ www.codexalimentarius.net



Sea scouts plan big splash for oceanography

Virginia Gewin, Washington

US researchers are drawing up a wish-list of requirements for an Integrated Ocean Observing System (IOOS), which they hope will radically strengthen the tools available for studying the world's oceans.

A group of 110 academics, government scientists and other researchers met in Warrenton, Virginia, earlier this month to hammer out a consensus on their priorities for the system.

The project has the firm backing of senior members of the Bush administration. And, because their input was requested by a Senate committee, oceanographers are optimistic that funds — perhaps several hundred million dollars in the first few years — will be made available. “The IOOS is an idea whose time has come,” says Lou Codispoti, an Arctic Ocean researcher at the University of Maryland's Horn Point Laboratory.

Researchers hope that, if approved, the IOOS will serve as a reliable, comprehensive source of ocean data, similar to the existing National Weather Service. Information would be gathered from buoys, satellites and ships, and would be available to everyone, from military planners to marine biologists. The data obtained would be relevant to issues such as climate change, forecasts of coastal conditions, and the status of fisheries.

The IOOS would have two components, dealing with the open ocean and with US coastal regions, respectively. It would also serve as the US component of the Global Ocean Observing System, an international project that is currently being implemented.

“We are in the midst of a real revolution of how oceanography is conducted,” explains Tom Malone, co-chair of the US Global Ocean Observing System Steering Committee. Malone is also joint leader of the IOOS planning efforts.

Advocates of the system say that it will help to transform oceanography, allowing measurements to be repeated easily over time. “The fundamental problem we face right now,” says Malone, “is that each piece of the existing system was developed by a particular agency for a particular purpose, so there's an awful lot of redundancy and at the same time we're not getting the information we need.”

Participants at the meeting discussed the limitations of current data networks, how new technology might be incorporated, and the logistics of putting new systems in place. A topic of particular interest, say meeting attendees, was the question of how to maintain and disseminate the volume of data that the IOOS is likely to collect. Another priority would be incorporating biological and chemical sensors into existing networks of buoys, built primarily for physical oceanography.

The conclusions of the workshop will be fed into a larger, interagency report to be sent to Congress this summer. This report will include cost estimates for the system, which Malone and others have so far declined to specify.

The US Commission on Ocean Policy, established early last year, is also expected to report in the summer on US national ocean strategy.

The commission's chair is Admiral James Watkins, a former president of the Consortium for Oceanographic Research and Education (CORE), the main lobby group for ocean research. Watkins is publicly

committed to the creation of the IOOS. Another advocate is Admiral Conrad Lautenbacher, administrator of the National Oceanic and Atmospheric Administration (NOAA), the US government's main civilian ocean research agency. Lautenbacher says that “there is no achievement too great to imagine from such a system being in place and working”.

Academic scientists have one big concern about the system — that its creation will eat into their grant funding from the National Science Foundation and other research agencies. “We're talking about making the case for new money,” says Malone. ■

Jospin reaches out to researchers



Looking to youth: presidential hopeful Lionel Jospin meets scientists during his visit to Genopole.

Sally Goodman, Paris

Lionel Jospin, the French socialist prime minister, has pledged a new deal for young researchers if he is elected president.

Speaking at Genopole, the science park in Evry, near Paris, on 13 March, Jospin promised to improve conditions for PhD students by providing an increased stipend and better career training. He also outlined a proposal to offer researchers permanent positions after two years of postdoctoral work, much earlier than is now the case.

“Increasing the number of posts is not enough to encourage young people to consider research careers if they have to spend years in postdoctoral positions,” Jospin said. He also promised to reduce the teaching loads of new university researchers, and to give bigger grants to teams with young leaders.

In addition, he pledged to reorganize the administration of research, creating a new ministry to deal with research, technology and higher education. Jospin had put

research under the control of a separate ministry after the sacking in March 2000 of education and research minister Claude Allègre, whose reform attempts sparked a rebellion in the research community (see *Nature* 404, 421; 2000).

Some researchers have still not forgiven Jospin for standing too long by Allègre, a geochemist at the Paris Geophysical Institute, who is still close to Jospin.

Jospin added that he would try to increase the amount of French gross domestic product that is spent on research and development to 3%, in line with a promise made by European leaders last week.

The two-round presidential elections, on 21 April and 5 May, pitch Jospin against Jacques Chirac, the conservative incumbent. Legislature elections follow in June, and could end a stand-off between Chirac and the legislature's socialist-dominated coalition. Observers say that a president and government of the same party would speed much-needed research reforms. ■

Dutch database aims to give voice to dying languages

Munich An ambitious attempt to create multimedia records of dying languages was launched last week by the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. Researchers plan to create a database containing sound and video recordings of rare languages being spoken.

The Volkswagen Foundation, a Hanover-based charity, has provided 3.5 million euros (US\$3.1 million) to fund 12 three-year projects. "Comparable information has only been available here and there, or in written form," says Martin Haspelmath, a linguist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.

Two-thirds of the world's 6,500 known languages are spoken by fewer than 5,000 people. For example, the Indian Wichita language in Colorado is now used by only 10 people.

♦ www.mpi.nl/DOBES

European satellite tracking system jostles for position

Barcelona Galileo, Europe's alternative to the Global Positioning System (GPS), has finally

been given the go-ahead by the European Union. The system, which will consist of 30 satellites, will primarily be used for transport and communications. Possible scientific applications include tectonics and climate research, flood management and long-term observation of animal movements.

The system will reduce Europe's reliance on the GPS, which is primarily run by the United States. Arguments over how to meet the project's costs — some 3.5 billion euros (US\$3.1 billion) — had threatened to derail the endeavour, but leaders of European Union's members reached agreement at a summit last week in Barcelona.

A joint European organization will oversee the construction of the satellites, which are due to be operational by 2008, in collaboration with the European Space Agency. The United States, which had criticized the project as superfluous, has nonetheless asked to participate in it to ensure that Galileo is compatible with the GPS.

♦ www.galileo-pgm.org

'Big dead lizard' is named and shamed

Washington Entomologist Michael Ivie says that his new name for the dinosaur formerly known as *Syntarsus* was meant as a joke.



Syntarsus, now known as *Megapnosaurus*: it's not big and it's not funny, say palaeontologists.

But his choice of *Megapnosaurus*, which means 'big dead lizard', has not amused palaeontologists.

Ivie, based at Montana State University in Bozeman, renamed *Syntarsus* after realizing that a beetle discovered in 1869 already had the same name. Following recognized taxonomy guidelines, Ivie issued a correction and attempted to inform the discoverer of *Syntarsus*, Mike Raath of the University of the Witwatersrand in Johannesburg.

But palaeontologists were not impressed. Some pointed out that the name is inaccurate, as *Syntarsus* was not big by dinosaur standards, and others have

disputed Ivie's right to rename it.

The controversy highlights problems with duplicate names. Thousands of biological species are thought to share the same names, and a recent paper (J. Alroy, *Proceedings of the National Academy of Sciences*; 10.1073/pnas.062691099; 2002) suggests that this and other taxonomy problems may have led to overestimates of global diversity.

Taste of success in search for oral smallpox drug

Prague Hopes for the development of an easy-to-take smallpox drug were set to receive a boost with the announcement by virologists of a treatment that can be taken orally.

Researchers at the International Conference on Antiviral Research in Prague were expected to reveal this week that hexadecyloxypropyl-cidofovir (HDP-CDV) protects mice against cowpox, a relative of smallpox. The drug also protects cultured human cells against smallpox.

The treatment has not yet been tested for safety in people, but John Huggins of the US Army Medical Research Institute of Infectious Diseases in Fort Detrick, Maryland, says he is optimistic that it will prove to be safe and effective in humans.

Once in the body, HDP-CDV is converted into the drug cidofovir, which is thought to be a safe smallpox treatment but which must be administered by injection.

Crop pioneer to lead non-nuclear family

New Delhi Indian geneticist Monkombu Sambasivan Swaminathan is to lead the Pugwash Conferences on Science and World Affairs, the influential nuclear non-proliferation group that was launched nearly 50 years ago by Albert Einstein and Bertrand Russell.

Swaminathan's work on crop genetics and sustainable agriculture has played an important role in helping to improve India's food supply over the past 40 years. He is currently chairman of the M. S. Swaminathan Research Foundation, a non-profit organization based in Chennai that works to promote sustainable development.

Swaminathan, who is the first Indian to lead Pugwash, was elected as its head last week at the 51st Pugwash Conference in Agra, India. The outgoing president, Michael Atiyah, said that Swaminathan's election would help to expand Pugwash's agenda into areas relating to health and welfare.

Germans up in arms over shotgun technique award



Craig Venter's win drew criticism.

Munich Some German scientists played party-pooper last week after Craig Venter, former president of Maryland-based Celera Genomics, won the prestigious Paul Ehrlich Prize. Venter won the award for developing the 'shotgun' method of mapping genetic sequences and

using it to map the human genome.

Helmut Blöcker of the German Research Centre for Biotechnology in Braunschweig, one of the German Human Genome Project's coordinators, told the German Press Agency that the award was "unjustified". He claims that the shotgun method is flawed and that the sequencing technique used by Venter was not new.

Many biologists feel that the sequencing of the human genome warrants a Nobel prize, but it is unclear whether both the public and the private projects will be so honoured. The Nobel committee will also have to decide how to deal with the rule that stipulates that a maximum of three people can receive the prize.

Harvard's melting pot

At a new genomics centre, ethologists are rubbing shoulders with computer scientists, chemists and mathematicians. Peter Aldhous visits a bold experiment in multidisciplinary.

Andrew Murray is used to standing out from the crowd. In the early 1990s, while rising through the academic ranks at the University of California, San Francisco (UCSF), his eye-catching array of facial piercings made him an instantly recognizable figure on campus.

More importantly, his elegant experiments into the checkpoints controlling cell division built his reputation as one of the leading biologists of his generation, and ensured that the brightest students beat a path to his door.

Now in his mid-40s, and having shed the facial metalwork, Murray is director of Harvard University's Bauer Center for Genomics Research, which celebrated its inauguration on 4 March. He is embarking on what could be his most ambitious experiment yet.

The Bauer centre aims to reap a post-genomic harvest by uniting physicists, mathematicians, chemists and computer scientists with a spectrum of biologists. They will work on collaborative projects dependent on high-throughput genomic analyses.

No one is quite sure what to expect yet, but there is a palpable sense of excitement among the centre's recruits. "This place, to me, looks like Disneyland," says mathematician Steve Altschuler.

As biologists struggle to make sense of reams of genomic data, many are wondering about enlisting the help of colleagues from other disciplines. But while most are still musing on the meaning of multidisciplinary, the Bauer centre is pressing ahead. Its strategy: hire a diverse and talented group of young scientists with a yen for collaboration, and throw them together. "We'll put them in



Shape of things to come: Doug Melton (below) persuaded Andrew Murray (right) to take on the role of director at the new Bauer centre (above).

a box and we'll shake really hard and hope that some fun things happen," says Murray.

Murray didn't actually plan to become the centre's director. He has always urged biologists to think in an integrated way, rather than simply filling in the molecular details in particular systems. But as the new millennium dawned, his growing interest in evolution was constrained at UCSF, which as a medical school could not indulge such diversions. So, in the summer of 2000, he moved to Harvard, setting up a lab to investigate evolutionary questions, in addition to his existing research in cell biology. One of his current projects, for instance, aims to simulate the origin of new species by applying pressures mimicking natural selection to laboratory populations of yeast.



Blazing a trail

As Murray arrived, the nascent Bauer centre was lacking a director. The original choice, Dari Shalon, a specialist in the production of DNA microarrays for analyses of gene expression, left when it became clear that his focus on technology development was at odds with the centre's wide-ranging mission. The centre's academic 'godfathers', developmental biologist Doug Melton and chemist Stuart Schreiber, who uses small molecules to disrupt gene function, set about convincing Murray to take over the reins. "After some trepidation, I said yes," says Murray.

Given the ambitions that Melton and Schreiber have for the Bauer centre, Murray's trepidation is understandable. Asked to



explain the centre's mission, Melton leaps to his feet and scribbles the word '*Leptothorax*', a genus of ant, on his blackboard. Melton's own research may be focused on stem cells within the pancreas, but he evidently has more than a passing interest in the biology of social insects.

Melton starts explaining the difference in morphology between different castes in *Leptothorax* ants. He then draws a squiggly path, as might be taken by a foraging worker. When a worker finds a food source, others are recruited and soon begin taking a more direct route. Melton chalks a straight line back to the colony's nest.

To understand *Leptothorax* biology, Melton argues, you must study the mechanisms of genome regulation that differentiate workers from other castes. You need expertise in behavioural biology, and the analytical skills to determine how creatures with simple nervous systems can collectively arrive at the solution to a complex navigational problem. Melton turns round from the blackboard:

"That's why we need the genome centre!"

The Bauer centre doesn't yet have a multi-disciplinary team working on the molecular mechanisms that underpin ant foraging, but in Hans Hofmann it has recruited a biologist interested in comparable conundrums.

Hofmann, a neuroethologist who joined the centre after completing a postdoc at Stanford University in California, studies cichlid fish from Lake Tanganyika in East Africa. These include a species called *Astatotilapia burtoni*, in which males can either defend nest sites to attract mates, or be non-territorial.

The two types of male look and behave very differently: territorial males are aggressive, have mature testes, and are yellow or blue with distinctive markings including dark stripes on their foreheads. Non-territorial males do not produce sperm, are sandy coloured, and school with groups of females.

Despite these profound differences, males switch readily between the two forms. At any one time, around a quarter of males are territorial, but there is a steady turnover — and any disturbance can overturn the *Astatotilapia* social order. For instance, if a hippopotamus churns up the lake bed, some territorial males will revert to the non-territorial form while previously non-territorial males seize their chance to reproduce.

Fish and chips

Hofmann wants to understand the shifts in gene expression that lie behind these changes, and how they are influenced by environmental and social cues. On each side of the *Astatotilapia* brain sit two populations of cells that secrete two hormones — gonadotropin-releasing hormone and somatostatin — known to be crucial to the shift to territoriality. Hofmann's studies involve dissecting out the tissue containing these cells, isolating messenger RNAs from the samples, and exposing them to DNA chips carrying the sequences of cichlid genes. By seeing where on the arrays the RNAs bind, he can tell which genes are active.

Working with the latest DNA microarray technology and expertise is crucial. But so, too, is the proximity of mathematicians such as Altschuler and his wife, Lani Wu. Hofmann suspects that making sense of his data will require a new conceptual framework that may only emerge from collaborations with people well versed in tough analytical problems. Altschuler and Wu's previous experience, for instance, includes working for Microsoft, developing algorithms for such tasks as separating background noise from audio signals — vital for the development of speech-recognition software.

Many former colleagues in behavioural biology remain flummoxed by Hofmann's residency at a genomics centre. But to him, this just shows that the gulfs between the various traditional disciplines within biology are greater than those dividing the diverse collection of scientists who have joined the Bauer centre. "Genomics, as I see it, will lead to a renaissance of organismal biology," says Hofmann. "There are still a lot of organismal biologists who do not see that and still a lot of molecular biologists who do not see that."

Some recruits have faced a steep learning curve. Altschuler and Wu have spent much of the past few months sitting alongside undergraduates in molecular biology lectures, and learning basic lab skills. They say this is crucial if they are not simply to become a service department for colleagues who want to analyse the function of networks of genes and proteins. "It's not just analysing the data; it's influencing the experiment upstream," says Wu. "Having your own bench space helps."

Battering ram

Altschuler says that a move into biology isn't for those who get their kicks by dazzling their peers with mathematical pyrotechnics. His and Wu's contribution may lie as much in introducing mathematical logic into the design and analysis of experiments as in devising sophisticated algorithms. "If we just contribute a couple of useful differential equations to a project, that's OK," says Altschuler.

Hofmann, Altschuler and Wu are just three of the centre's eight 'genome fellows', who were all selected for their willingness to collaborate and their multidisciplinary outlook. Laura Garwin, the centre's director of research affairs and formerly *Nature's* North American editor, describes the fellowship programme as a "battering ram" that will assault the walls dividing scientific disciplines. The fellows will be appointed for three to five years, each running a research group of up to three people.

Murray is delighted with the first crop of fellows. But recruitment remains a high pri-



Mathematicians Lani Wu and Steve Altschuler are adding a new dimension to the Bauer centre.

CHRIS SANDER

ority. Three vacancies remain, and Murray is especially keen to fill one of them with a theoretical physicist. "They have a relentless tendency to reduce problems to the fundamentals," he says.

Having got his cast of fellows together, Murray intends to take a hands-off approach to managing their work. He describes the Bauer centre as an "artists' colony", contrasting it with the "paramilitary organization" of explicitly goal-oriented genome centres. Nevertheless, there will be incentives to encourage fruitful interactions, including a pot of money specifically for collaborations between the fellows.

The Harvard authorities are betting on the Bauer centre spawning collaborations across the university more generally — and, in this spirit, the centre has opened its doors to Harvard scientists wanting to perform genomic analyses. The building into which the centre moved earlier this month was financed by a \$25-million donation from investment manager and Harvard alumnus Charles 'Ted' Bauer, but its initial \$6-million annual operating budget comes from the university's own reserves. It is part of a wider initiative to invest in new scientific developments, after concerns were raised about Harvard's future ability to compete at the very highest level.

Murray, who is aware that ambitious experiments can sometimes end in failure, has some reservations about the expectations being pinned on the centre. "There's a lot of hope being placed into what is still a fairly frail vessel," he says. But his new recruits are finding it hard to contain their enthusiasm. "It's the place of the future, we hope," says Hofmann. ■

Peter Aldhous is *Nature's* chief news and features editor.

▶ www.cgr.harvard.edu



Hans Hofmann hopes to discover why these male cichlid fish look and behave differently.

BAUER CENTER: MARTHA STEWART/C. RUSSELL FERNALD

Peer review, unmasked

The editorial review of scientific papers usually takes place behind closed doors, but could the process be improved by opening it up for all to see? Trisha Gura investigates.

You have written an interesting but provocative paper that is likely to stir up debate. Where should you try to publish it? Stefan Füglistaler, an atmospheric chemist at the ETH, the Swiss Federal Institute of Technology in Zurich, found himself asking this question last December. In the end, he chose not to submit his work to one of the field's established journals, but to an online newcomer with an unusual approach to peer review.

Füglistaler and his colleagues submitted their work—a new explanation for how large nitrogen-containing particles originate in the Arctic atmosphere—to *Atmospheric Chemistry and Physics* (ACP). Before sending papers out for peer review, ACP posts them online and holds a commentary session where scientists can debate the work, or simply offer helpful pointers. For new ideas such as Füglistaler's, it seems like the perfect testing ground.

ACP is not alone. A handful of other journals have launched experiments in 'open' peer review. Thanks to the Internet, the kind of debate that takes place at conferences can now be incorporated into the editorial process. For advocates of opening up peer review, it is an idea whose time has come. "If



Stefan Füglistaler: testing the air.

I'm right," says Drummond Rennie, a deputy editor of *The Journal of the American Medical Association* (JAMA), "all journals will be doing this in the future."

The founders of ACP believe that they are providing an alternative to a flawed system. In traditional peer review, editors solicit comments on a paper from relevant experts, and use these as

a basis for a decision on whether to publish. Most scientists agree that the process, if performed properly, is a good way of assessing new research. But very few would argue that it is problem-free.

Authors sometimes claim, for example, that good work is rejected because it clashes with the reviewers' own studies or opinions—or simply because the ideas expressed are too 'left field'. Reviewers can also miss technical errors. "I sometimes wonder how some papers get through," says Thomas Koop, an executive editor of ACP, who is also at the ETH. Papers that span different disciplines cause particular problems, as individual reviewers are often only familiar with one of the fields involved.

Climate of change

These issues got atmospheric scientists talking. In June 2000, Ulrich Pöschl, an atmospheric chemist at the Technical University of Munich, approached his fellow researchers with a proposal to develop a forum where papers could be discussed before being submitted for formal peer review. Scientists in the field would have a chance to offer tips and comments, and the authors could then defend or revise their work.

The group decided to launch an online journal with open commentary sessions prior to formal peer review, and convinced

the European Geophysical Society to host the experiment. The founders formed an editorial board, designed a website and laid out the details of how the commentary stage would work. Last September, ACP was born.

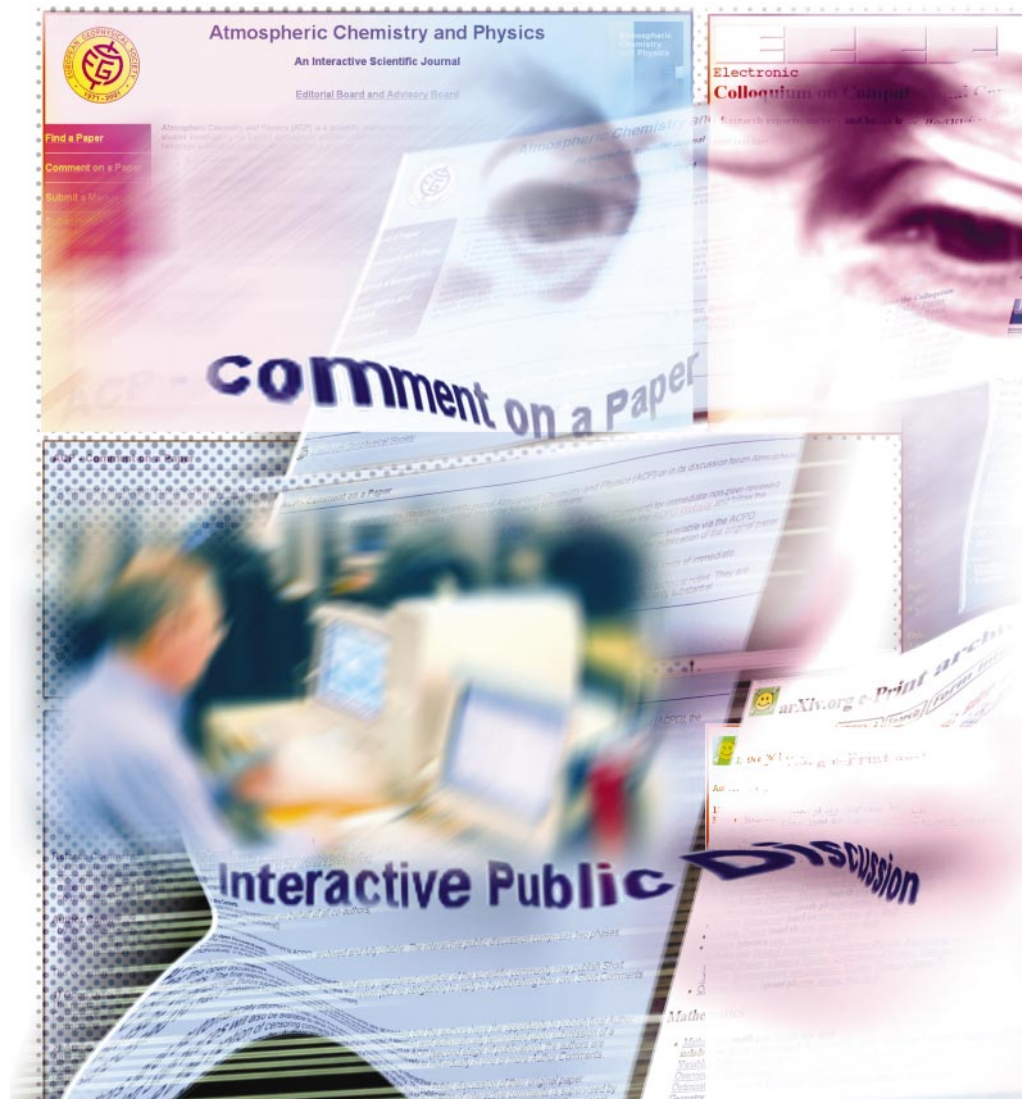
The submission process starts with a quick once-over from relevant members of ACP's 60-strong editorial board. An editor assigned to the paper decides whether technical corrections are needed. If the paper meets the basic standards, it is posted on ACP's website. Registered researchers can then post comments, to which the authors can respond.

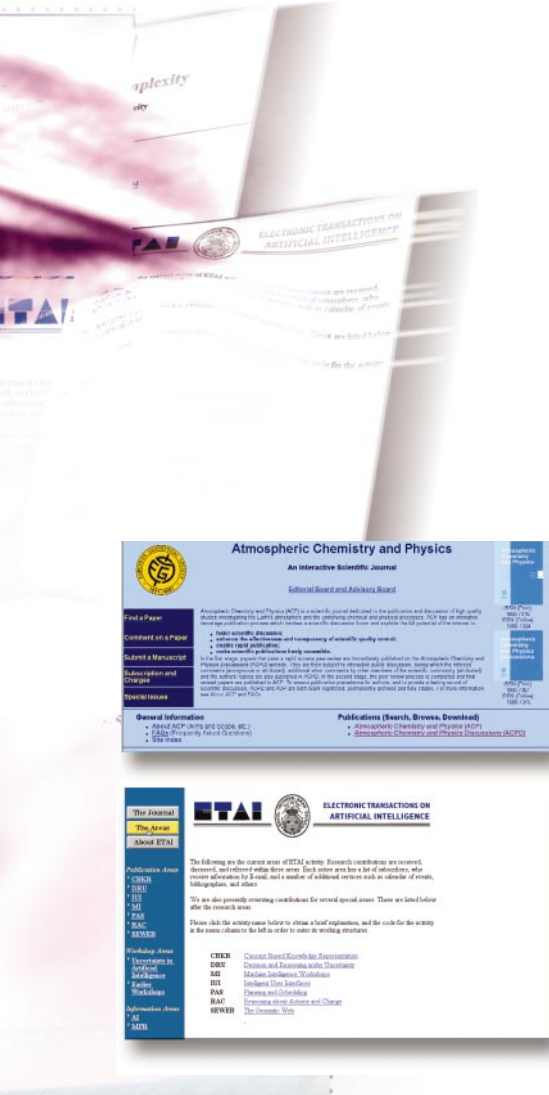
The discussion is moderated by the assigned editor, who can edit out any personal attacks or inflammatory comments. At the end of eight weeks, the authors have



Drummond Rennie: ready for change.

the option of revising the paper or submitting it for traditional peer review. The reviewers for this latter stage are selected before the paper goes online—and they can also comment during the initial stage, albeit anonymously. If a manuscript makes it through the entire process, as 12 have done to date, it is officially accepted for publica-





received only three comments, two from the reviewers assigned to the manuscript, was disappointed with the lack of discussion.

"It will take a little time to get acquainted with the idea of interactive commenting," says Pöschl. But he hopes to see an increase in activity now that the European Geophysical Society has started advertising *ACP*.

Although *ACP* has yet to establish itself, successful experiments in other fields suggest that the process can work. After a debate at the 1996 European Conference on Artificial Intelligence, held in Budapest, researchers in the field asked Erik Sandewall, a mathematician and computer scientist at Linköping University in Sweden, to develop a new journal that would make peer review more discursive. "Criticism at seminars is seen as valuable to scientists," says Sandewall. "If there are going to be critical comments, it is better to capture them at an early stage."

Six months after the meeting, *Electronic Transactions on Artificial Intelligence (ETAI)* started recruiting editors. Like *ACP*, *ETAI* relies on open discussion moderated by editors, followed by confidential review. But there are slight differences between the two.

Artificial intelligence (AI) is made up of a relatively small number of related fields, and *ETAI* is divided into sections that cater for these subdisciplines. Subscribers are alerted by e-mail when a paper in their area is posted on the website. Signed comments from researchers, as well as anonymous notes from preselected reviewers, are e-mailed to all subscribers as well being archived in an online log linked to the paper. At the end of the discussion, authors can revise their manuscripts. The paper then goes to the reviewers, who have already asked their questions and now just stamp the paper as accepted or rejected.

Rapid results

Iliano Cervesato, a computer scientist at ITT Industries in Alexandria, Virginia, believes that *ETAI* offers authors better insights compared with traditional journals. And because it is not hampered by print production schedules, it can publish papers more quickly despite the additional open commentary step. *ETAI* published a paper from Cervesato's group in just six months — a sprint compared with the one to two years that AI journals often take, he says.

AI researchers normally present fresh ideas at meetings in the form of technical reports. Although not "masterpieces of literature," says Cervesato, these reports mark a researcher's territory during the long wait for formal publication. *ETAI* offers a halfway house, helping researchers get peer-reviewed and better-written versions of technical reports into the public domain quickly.

Whether or not open peer review really offers any advantages over traditional techniques should become clearer when *The Medical Journal of Australia (MJA)* finishes

the second of its peer-review experiments. In the first study, which started in March 1996 and lasted for just over a year, papers that had already been accepted for publication were posted online, together with comments made by referees during the review process. Readers were invited to discuss the papers, and authors had the option of revising their papers before publication in the print version after seeing what the readers had to say.

Around half of the 56 papers attracted comments, and about 2% of *MJA*'s readers took part in the debates. Seven authors revised their manuscripts as a result. The journal's editors concluded that open peer review is useful, but not a substitute for formal review.

Paper chains

The journal has since begun a larger study. Starting in October 1998, some papers have been posted on a password-protected website. The review consists of an online discussion to which authors, editors and assigned peer reviewers have access. A panel of six consultants, chosen to represent a broader range of the journal's readership than the peer reviewers provide, can also comment.

Editors decide whether to accept, reject or ask for changes after three to four weeks of discussion. Accepted papers, and the associated discussions, are then placed on an open website and readers are invited to comment. Authors can again revise their papers on the basis of these comments before publication in the print version.

The editors running the study are rating the quality and efficiency of the procedure by discussing it with authors and reviewers, and comparing the results with those from a control group of papers that have gone through traditional peer review.

The trial is currently on hold while *MJA* upgrades the website it uses for the debates, although Craig Bingham, the journal's communications development manager, says that he is confident the system will improve dialogue between authors and reviewers. The *British Medical Journal* is conducting its own study of open peer review but is reluctant to discuss the experiment as the editors want to submit the results for publication.

But not everyone is impressed with open peer review. Stevan Harnad, a cognitive scientist based at the University of Southampton, UK, has spent 25 years editing the journal *Behavioral and Brain Sciences*. An editor, he says, is a gatekeeper. He points to the papers that pass across his desk. "As an editor, I waste a lot of time on raw sludge," he says. "We'd waste even more time if everyone looked at papers online without knowing which ones were good and which were bad."

Advocates of open peer review counter that the process need not entail sacrificing quality control. Editors can still weed out technically deficient papers, they argue. And the nature of an open forum, they claim,

tion. And, because *ACP* is an online journal, papers can be made available immediately.

Kaarle Hämeri, an atmospheric physicist at the Finnish Institute of Occupational Health in Helsinki, has had two papers accepted for publication by *ACP*. He says that *ACP* published his work faster than other journals he had submitted to, and rates the quality of the other papers published as high. But, like other authors who have published with *ACP*, Hämeri found his paper attracted only limited debate. So far, it seems, the atmospheric-science community hasn't exactly jumped at the concept of open peer review.

Füglister's paper, for example, generated just two online comments, before moving on to formal review. Hämeri, whose first paper



Peers go public: Thomas Koop (right) and Ulrich Pöschl back the idea of open commentary.

► motivates researchers to put up better papers. "There is an embarrassment factor," says Sandewall. "People think twice about submitting garbage."

David Poole, a computer scientist at the University of British Columbia in Vancouver, agrees. He published a paper on decision theory and logic in *ETAI* in 1998. In order to brave the threat of peer attack, he needed to be confident about his work. "You can't go on a fishing expedition," he says.

Harnad accepts that open peer review might work well within small, cooperative communities such as the AI fraternity, but argues that it is difficult to implement for journals with broader readerships. As an alternative, Harnad points to online archives, where papers can be deposited before or at the same time as they are submitted to journals. Many such archives have sprung up in the past 10 years, and some scientists see communication between the researchers who use them as an informal form of peer review.

Into the vault

Some archives, such as the Electronic Colloquium on Computational Complexity, which started storing papers in this mathematical subject in 1994, involve a small amount of editorial input. The colloquium has an editorial board, but members merely check papers for technical errors and ensure that the submission is correctly categorized. Comments and corrections are kept with the original paper, but there is no final acceptance process and authors often submit their paper to a traditional journal after it has been discussed online.

Other archives have even less of a role for editors. The arXiv server housed at Cornell University in Ithaca, New York, and founded 10 years ago by physicist Paul Ginsparg at the Los Alamos National Laboratory in New Mexico, is the best-known example. The archive now allows researchers to post physics, computer science and mathematics papers online before they are submitted for publication, and imposes no editorial control.

Harnad, who runs CogPrints, a comparable archive for the cognitive sciences, says that most submitted papers will eventually go through conventional peer review, and appear in a traditional journal. Most archive users seem happy that these repositories coexist alongside traditional journals, but a minority



argue that the informal peer communication provided by archives will eventually replace formal publication. "By the time I go through peer review, my work is old news," says theoretical physicist Giulio Ruffini, director of Starlab Barcelona, a private research centre that places little importance on publishing in peer-reviewed journals. "Archiving is fast and easy and puts a time stamp on what you have done so that you can later claim ownership."

Stolen words

But in other fields, archives have not taken off. In medicine, for example, some researchers have argued that uncorrected mistakes in a manuscript could cost lives. And in molecular biology, some papers are like recipes: having read the methods section, it would be easy to repeat the work and publish it elsewhere.

The potential for plagiarism in fast-moving, highly competitive fields may also limit the adoption of open peer review. Already, journal editors deal with occasional accusations of plagiarism by reviewers (see *Nature* **413**, 102–104; 2001). Authors also frequently name individuals whom they would like excluded as reviewers — just imagine the flood of complaints if all researchers routinely had access to molecular biology papers being discussed online prior to formal publication.

Such objections might explain why few of



Paul Ginsparg (above) has recently wheeled the successful pre-print archive he set up to a new home at Cornell University (left).

the world's 20,000 or so peer-reviewed scientific journals are flirting with open peer review. Those that have, such as *ACP*, tend to be new or less prestigious publications.

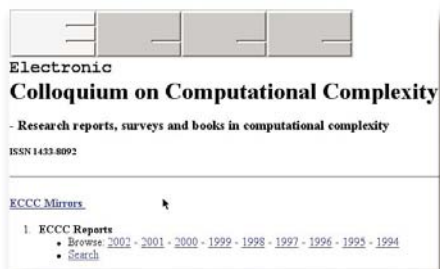
Philip Campbell, editor of *Nature*, says that some of the innovations are worth watching. But neither *Nature* nor *Science* has plans to implement any form of open peer review. And, despite Rennie's enthusiasm, *JAMA* is not planning to change the way it handles papers either.

Part of the problem, says *Science*'s editor-in-chief, Donald Kennedy, is a lack of information about alternative methods of handling submitted papers. "What we have now is a situation with lots of models out there and no systematic effort to bring them together and look at the advantages and disadvantages of each," he says. For smaller journals, the logistical difficulty of moderating an online commentary poses an additional problem.

Rennie — a long-standing proponent of innovation in peer review — argues that cultural inertia is also a factor. "There is an inborn conservatism of scientists, which is huge," he says. Andrew Odlyzko, a mathematician at the University of Minnesota in Minneapolis who studies trends in online publication, agrees. "Change is usually slow, especially when it is sociological change," he says.

Without a clear demonstration that open peer review makes for better papers, few journals seem likely to adopt new methods. But for advocates of reform, the opportunities for innovation allowed by online scientific discussion are too great to ignore. Changing peer review, says Rennie, is "one grand experiment".

Trisha Gura is a freelance writer in Cleveland, Ohio.



Web of words: online archives can allow informal peer review of papers before formal publication.

Biotech remains unloved by the more informed

The media may be providing the message — but is anyone heeding the call?

Sir — Public hostility towards biotechnologies is frequently attributed to lack of information, due to poor and insufficient media coverage. For this reason, scientific researchers and policy-makers often call for journalists to give more attention to scientific issues, for better information campaigns and for more communication of science, to improve general understanding and thereby lead to greater public support for biotechnologies and other innovations. But is this approach correct?

In 2000 and 2001, with partial support from the Giannino Bassetti Foundation, we carried out two surveys of Italian public opinion. These were specifically to analyse the relationships between exposure to science in the media, information on biotechnologies, trust in science, and attitudes to biotechnologies. A representative sample of 1,022 Italian citizens aged over 18 were interviewed by phone in September 2000; another representative sample of 1,017 citizens were interviewed in November 2001. Some questions were identical for the two groups, others were year-specific. (A copy of the full list of questions used in the survey and the percentage response rates is available from M.B.)

Respondents were asked about their level of exposure to science in specified daily newspapers, television and radio science programmes, popular science books and magazines. We used questions similar to those of 1999 Eurobarometer (see <http://europa.eu.int/comm/research/pdf/eurobarometer-en.pdf>), but also asked additional ones about trust in science and scientists, and the use, risks and moral acceptability of biotechnologies.

Our results confirm previous suspicions that exposure to information does not always lead to greater trust in biotechnologies. We also find that greater exposure to science in the media does not necessarily mean a higher level of understanding. The proportion of subjects who think “only genetically modified tomatoes contain genes while ordinary tomatoes don’t”, for example, is almost identical among those with high (29%) and low (31%) exposure to science in the media. More than a quarter of the ‘regular’ consumers of science in the media (28%) cannot give more than one correct answer to five questions about biotechnologies, and more than half (57%) cannot give more than two correct answers.

High exposure to science in the media does not significantly reduce opposition to

applications such as “taking genes from plant species and transferring them into crop plants, to make them more resistant to insect pests” or “introducing human genes into animals to produce organs for human transplants, such as into pigs for human heart transplants”. But it does result in greater criticism for some applications: 64% of the most exposed subjects consider embryo research to be ethically unacceptable compared with 59% of the less exposed, and 80% of regular consumers of science in the media consider reproductive cloning useless compared with 76% of low consumers.

Of course, media exposure to science does not guarantee accurate information; indeed, there are frequent complaints about the quality of science coverage by the mass media. People who are exposed to at least one high-quality source of public communication of science (for example, the Italian edition of *Scientific American*) are more likely to have a positive attitude to biotechnologies. Yet this result merely highlights a well-known paradox in the communication of science: the greatest impact is on a small minority, who are most likely to have the information already.

A high level of information does not guarantee a positive attitude: 49% of the better-informed respondents think that transferring genes into fruit or vegetables is useless, and 54% think it is risky. Embryo research fares poorly (60% in both groups consider it unacceptable),

whereas cloning for reproductive purposes is even more severely judged by the better informed than by the less well informed.

A higher level of information is associated with the desire for stricter state regulation of biotechnologies, as well as with the belief that regulation should not be left either to companies or to scientists alone. The better informed are also more likely to trust consumers’ organizations and scientific institutions more than potential beneficiaries (such as patients’ groups) and, sometimes, government institutions.

If media exposure to science does not account for different attitudes to biotechnologies, what does? Attitudes appear to be rooted at a deeper, cultural level where values (such as trust and conception of risk) are heavily involved and media information does not reach. Public awareness of biotechnologies is increasing and the level of education seems to be more important than other factors in explaining attitudes in this area. So it may be wise to recommend that at least as much attention is devoted to science education — both in terms of research and of programmes and investments — as to the mass-media communication of science.

Massimiliano Bucchi*, Federico Neresini†

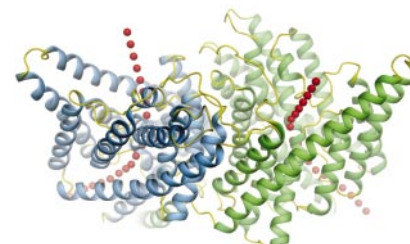
*Department of Social Sciences, University of Trento, via Verdi 26 - 38100, Trento, Italy
e-mail: mbucchi@soc.unitn.it

†Department of Sociology, University of Padova, via S. Canziano 8, 35122 Padova, Italy

Nothing automatic about ion-channel structures

Sir — My colleagues and I were shocked to read your News report “Protein chemists favour automatic answers” (ref. 1) in which the chloride ion channel was featured prominently as an example of an important protein structure determined with the help of high-throughput techniques. In the report, Neil Isaacs of Glasgow University is quoted as saying that the chloride ion-channel structure “could not have been done without automation”.

In fact, we used no automation or high-throughput methods to solve the chloride-channel structure². Indeed, high-throughput methods have played no part in any of the difficult ion-channel structure determinations completed in my laboratory^{3–5}. Our success has rested solely



Chloride ion channel: structure solved by traditional science.

on the intense focus, hard work and thoughtful approach of a small group of scientists intent on solving an important problem in biological chemistry.

I do not wish to join the debate over the wisdom of funding robotic structural biology in the United Kingdom. I do, however, wish to set the record straight concerning a misrepresentation of the science carried out in my own laboratory. The explanation for why we have made

good progress is simple — we try hard to understand the proteins that we work with. We study the structure and function of ion channels in one small laboratory. I do not have a ‘structure group’ and a ‘function group’; young scientists who come to work with me pursue the structural and functional sides of their projects as a single endeavour. If you want to solve the structure of an ion channel it helps to understand ion channels. To understand ion channels you must study their function. Our approach is not profound, it is traditional hypothesis-driven science, and it works.

Roderick MacKinnon

Howard Hughes Medical Institute, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

1. Adam, D. *Nature* **415**, 822 (2002).
2. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. *Nature* **415**, 287–294 (2002).
3. Doyle, D. A. et al. *Science* **280**, 69–77 (1998).
4. Morais-Cabral, J. H., Zhou, Y. & MacKinnon, R. *Nature* **414**, 37–42 (2001).
5. Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. *Nature* **414**, 43–48 (2001).

Opportunities for women in science (Russia, 1912)

Sir — Caroline Herzenberg (*Nature* **414**, 843; 2001) is correct to suggest in Correspondence that the two women in the photograph of Ivan Petrovich Pavlov and his group are Maria Kapitonovna Petrova and Maria Nikolaevna Erofeeva (alternatively spelt “Nikolayevna Yerofeyeva”). The original, taken in Pavlov’s Department of Physiology at the Military Medical Academy in 1912, includes Pavlov, seventeen male co-workers and the two female co-workers.

Erofeeva and Petrova were members of the first generation of women to enter Pavlov’s laboratory, capitalizing on the expanded opportunities for women in the wake of the 1905 revolution. Erofeeva performed important experiments in 1910–12 that supported Pavlov’s view that any environmental agent could, in principle, become a conditional stimulus. She defended her doctoral thesis in 1912 and, unlike most of Pavlov’s co-workers, continued to work in his laboratory until taking a position at the Petropavlovsk Hospital and the P. F. Lesgaft scientific institute, where she continued her research until her death in 1925.

Petrova completed her doctoral thesis, on the nature and interaction of excitation and inhibition, in 1914. She and Pavlov became lovers that same year, Petrova becoming Pavlov’s most important

co-worker from about 1921 until his death in 1936. She created and was the principal force behind the laboratory’s increasingly important line of investigation into “the experimental pathology and therapeutics of higher nervous activity”. She directed the department of physiology and pathophysiology of higher nervous activity at the Leningrad Institute for the Improvement of Physicians from 1935 to 1941, and was also a prolific researcher and laboratory director at the Physiological Institute of the USSR Academy of Sciences. She survived the siege of Leningrad, won a Stalin prize for scientific research in 1946, and died in 1948.

Pavlov himself conducted very few of the experiments underlying his Nobel Prize-winning research on digestion and his more famous investigations on conditional reflexes, so his co-workers were central to his laboratory’s research — a subject I have previously covered in *Pavlov’s Physiology Factory: Experiment, Interpretation, Laboratory Enterprise* (Johns Hopkins University Press, Baltimore and London, 2002). I hope to rescue Erofeeva, Petrova and other equally interesting co-workers of the later period from obscurity in my forthcoming biography of Pavlov.

Daniel P. Todes

Department of History of Science, Medicine and Technology, The Johns Hopkins University, 1900 East Monument Street, Baltimore, Maryland 21205, USA

Job-seekers, be careful of what you’re signing

Sir — My colleagues and I now strongly advise all our undergraduate and graduate students planning to enter industry to “read the fine print” — specifically, any exclusion clauses. It pays to check out this apparently small but possibly very significant detail before firmly accepting an offer, certainly before signing a contract or relocating. We feel that this precaution is valuable to any professional considering a change of employer, in addition to the familiar considerations of job satisfaction (of prime importance); type of work; size, type and health of company; location; and company benefits.

Some companies ask new employees to sign an agreement covering issues of patent ownership and confidentiality, for example agreeing that the company is the owner of any technical idea developed by the employee; or that the employee must not disclose or use outside the company any information that is proprietary to the

company, either while an employee or afterwards. Such agreements, which must be signed upon starting a new job, are commonplace to cover the legitimate right of the company to protect its proprietary ideas and technology.

But some companies attempt to impose other, considerably more demanding, requirements. One might be that a new employee agrees, for a stipulated time after taking employment with a different company, not to work in areas of research or technology in which he or she had been directly involved while working for the present company. A more extreme requirement is that the employee is not allowed to work or consult for a competitor for a stated period of time after leaving the company, sometimes for as long as two years.

Typically, of course, a person accepting a position with a new company has no thought of working elsewhere. But who can count on anything lasting forever? The level of job satisfaction could change over time; the employee may wish to move; the work environment might become degraded through economies or ineffective management; staff could be laid off, as happens often in industry these days; or the employee might simply find a new opportunity with another company. Such reasons for change can arise relatively early in a person’s career or after many years.

When embarking on a possible career move, the employee may have forgotten signing the agreement, but the company management will remember, and will remind the employee of the binding nature and implications of agreements that operate beyond the term of employment with the new employer.

So, our core advice to our students is that — before they make their final decision among different job offers, and before they have left their university town or home town — they ask any company offering them a job to send them copies of all agreements that they will be asked to sign when they join the organization. They can then seek legal advice in case of troubling ‘non-compete’ clauses. The requirements written into confidentiality agreements range widely from the quite mild and reasonably undemanding to those that go beyond the bounds of what I personally think ethically justifiable and proper.

Whatever the specifics, all these agreements are written in the interest of giving the company the protection that it feels it needs, and not for the benefit of the individual employee.

Ken Lerner

Department of Geophysics, Colorado School of Mines, Golden, Colorado 80401, USA

Unsung heroes of the revolution

The people whose ideas laid the foundations for the information age.

The Chip: How Two Americans Invented the Microchip and Launched a Revolution

by T. R. Reid

Random House: 2001. 309 pp. \$13.95, £9.85

Go To: The Story of the Math Majors, Bridge Players, Engineers, Chess Wizards, Scientists and Iconoclasts Who Were the Hero Programmers of the Software Revolution

by Steve Lohr

Basic Books: 2001. 250 pp. \$27.50, £19.99

Paul Peercy

Two intertwined technologies, the integrated circuit and software, underlie today's high-technology world. The integrated circuit, or 'chip', transformed semiconductor technology from discrete transistors to monolithic circuits that today contain tens or hundreds of millions of transistors, along with associated electrical components, in a single piece of silicon. Software programs enable the use of this 'hardware'. These technologies have led to the so-called 'information age' and have fundamentally changed the way people around the world work, play and live. *The Chip* by T. R. Reid and *Go To* by Steve Lohr both present a fascinating overview of the challenges faced by the inventors of these technologies and how they arrived at the solutions.

Intrigued by how few people recognized the names Jack Kilby and Robert Noyce, who invented the integrated circuit, Reid decided to write the story of the scientists and engineers behind the semiconductor revolution.

Kilby, while an engineer at the electronics company Centralab, intuitively recognized the potential impact of the transistor and led his colleagues in helping develop the transistor into mainstream electronics. It was here that Kilby encountered the 'tyranny of numbers' — engineers could design circuits with tens of thousands of transistors, along with the other required electrical components, but it was virtually impossible to hook all these components together without error. Reid traces Kilby's move to Texas Instruments to pursue his quest to solve this problem, the scepticism of his engineering colleagues there, and the success in 1958 of his 'monolithic idea'. This costly work was funded by the US government, which was seeking lightweight, low-power-consumption electronics for space and defence applications.

Noyce, a co-founder of Intel, was also looking for a solution to the 'tyranny of numbers', and in 1959 he independently had the idea of building the entire circuit



Computer pioneer Jack Kilby ponders a future of integrated circuits made on 300-millimetre wafers.

in silicon. Many of the techniques used in Noyce's approach were forerunners of high-volume manufacturing techniques.

Equally interesting is the account of how Japan developed a formidable microelectronics industry that threatened to destroy the US semiconductor industry. The US Semiconductor Industry Association was formed to voice concerns about the potential consequences to national defence and the economy if this industry were to be lost, and a government-industry partnership called SEMATECH was set up in 1987 to help the US semiconductor industry recover.

Reid covers much more than the invention of the integrated circuit. He traces the history of electronics, beginning with the physics that was to lay the groundwork for the vacuum tube. Later, managers and scientists at Bell Laboratories recognized the need for a solid-state replacement for the vacuum tube, which led to the invention

of the transistor. *The Chip* is a well-written book that is worth reading.

Go To is a *tour de force* in which Steve Lohr recounts the history of software, from the earliest days of the original computers, which were programmed by setting switches in the hardware, to today's World Wide Web. In the early days, the task of machine-language programming was carried out by a select group of specialists, and it could take days or weeks before the computer was 'programmed' to calculate the solution to a technical problem.

The name of the book is taken from one of the commands in the software language Fortran. ('Go to' instructs the computer to jump to any place in the program that the programmer chooses.) Created by a small group at IBM, Fortran (from formula translation) was the first software language to move programming from machine language closer to the language of humans. Despite the programming community's conviction

that software could not be created that would run a program as efficiently as the tedious machine-language programs, this eclectic group proved the doubters wrong. Programs in Fortran not only ran efficiently, but also changed the way people thought about software and made the computer much more widely accessible.

Lohr recounts the numerous computer languages created after Fortran, along with the intriguing personalities of their creators. He takes the reader on a fascinating tour through Cobol to Unix and C, through BASIC and Visual BASIC, and Algol to Pascal and C++. By this time, the stage is set for the personal computer and the rise of software as a separate business, and the end of the mainframe era when software was shipped as part of the computer. Although we have seen many advances that significantly increased the utility of the computer, software development continues to lag behind hardware and limits the performance and use of cutting-edge computing.

Lohr ends with the creation of the markup language HTML for the web and the philosophical battle being waged over open-source software such as Linux, for which the source code is freely available, and commercial software. It will be fascinating to see how the

battle plays out over the next decade or so. ■
Paul Peercy is at the University of Wisconsin, 2610 Engineering Hall, 1415 Engineering Drive, Madison, Wisconsin 53706-1691, USA.

In the eye of the beholder

It Must Be Beautiful: Great Equations of Modern Science

edited by Graham Farmelo

Granta: 2002. 224 pp. £20, \$25

Ian Stewart

When Robert Recorde invented the equals sign he justified his choice of two parallel lines on the grounds that “no two things can be moare equalle”. His inspired symbol, now indispensable in science and mathematics, forms an iconic decoration for the dustjacket of *It Must Be Beautiful*.

The beauty of equations passes most of the human race by. But now we have a brilliantly readable book that absolutely revels in that beauty. Graham Farmelo has assembled a star-studded cast of authors and encouraged them to wax lyrical about their favourite

equations. The book's main message is the one that Paul Dirac scrawled on a Moscow blackboard in 1955: “Physical laws should have mathematical beauty.”

The essays are arranged in chronological order, based on the date of origin of the equation. Farmelo kicks off with the Planck–Einstein equation for the energy of a quantum of radiation, the discovery that started the quantum revolution. His essay is mainly historical, and the characters come to life before our eyes. Einstein turns up in several essays, probably because he set himself such high standards. We are told that the poet Paul Valéry kept a notebook to jot down his ideas, and asked Einstein if he did the same. “Oh, that's not necessary,” said Albert. Then he added wistfully: “It's so seldom I have one.” The ideas he did have, though, were stunning. It falls to Peter Galison to remind us of Einstein's most famous equation. It is, of course, $E = mc^2$. That's how Einstein first wrote it, if it seems unfamiliar.

Physics takes up about half of the book. Roger Penrose makes a valiant effort to explain general relativity, and tells us that the total energy of the gravitational waves radiated by Jupiter is about that of a 40-watt light bulb. Arthur Miller dishes the dirt on Schrödinger's equation — and his love life. Frank Wilczek revels in Dirac's relativistic equation for the electron, and Christine Sutton does a wonderful job on the Yang–Mills equations. The rest of the book is more varied: Igor Aleksander on Shannon's equations for information, Robert May on the logistic map of chaos theory, and John Maynard Smith on evolution. All of the essays are excellent, but some are more excellent than others.

Oliver Morton's contribution is rather offbeat: the Drake equation, used to estimate the prospects of finding intelligent aliens. As Jack Cohen and I point out in our forthcoming book *Evolving the Alien* (Ebury Press), the equation itself is open to serious criticism — as, even more so, is its extension in Peter Ward and Donald Brownlee's *Rare Earth* (Copernicus, 2000). But its history is fascinating. Even more fascinating is Aisling Irwin's account of the Molina–Rowland equations, which explain why chlorofluorocarbons destroy ozone, and predicted that they would. Those equations may just have saved the planet.

Finally, Steven Weinberg muses wisely on the staying power of a good equation, pointing out that all of the equations in the book are actually wrong — and are all the better for it. A comprehensible and insightful model beats an exact description any day.

It would have rounded off the discussion if at least one purely mathematical equation had been included — maybe Euler's $e^{i\pi} = -1$, or one of Ramanujan's partition formulas. The beauty of equations resides in their mathematical form and implications; applications relate to their importance as models of the real world. But the two are linked,



A Renaissance world view

The Nuremberg Chronicle, more correctly known as the *Weltchronik*, or *Chronicle of the World*, was an extraordinary Renaissance achievement, and one of the most important works ever published. It is an illustrated encyclopedia of the world as it was known in Germany near the end of the fifteenth century. Most of its knowledge is drawn from the Bible, but it also records contemporary events, including the falling of a meteorite on the Alsace, and some history of medicine. It includes

maps of important cities, and a map of the world — just too early to include Columbus's new findings. Compiled by Hartmann Schedel in Nuremberg, the book was illustrated by renowned artists and engravers including Albrecht Dürer, who went on to become Germany's best-known Renaissance painter. Taschen has now published a luxurious facsimile of the massive chronicle (£40, \$60, 60 euros), with extensive commentary in several languages.

which was Dirac's point. Many of the most beautiful equations of mathematics originated in questions posed by science. But some beautiful equations didn't, and that ought to be said, too. Let's hope there is a sequel. ■

Ian Stewart is in the Department of Mathematics, University of Houston, Houston, Texas 77204-3476, USA.

A milestone for a new millennium

The Human Genome

edited by Carina Dennis & Richard Gallagher

Palgrave: 2001. 156 pp, £19.99, \$30

Daniel Cohen

The sequencing of the human genome was the last major goal to be set by leaders of science in the twentieth century. Its achievement was certainly as spectacular a triumph as the launching of the first spacecraft or landing on the Moon. It was also an amazing example of cooperation and competition between scientists, government funding agencies, charities, and public and privately funded projects. In a symbolic way it is a manifestation of the dramatic cultural changes that humanity undergoes at the turn of a new millennium.

Nature was an active supporter of the project to sequence the human genome, promoting ideas, lobbying for more funds and publishing outstanding achievements, including a completed draft sequence. So it is natural to see two senior editors of *Nature* compiling a book on this epochal achievement. The book is aimed at both non-specialists and students in the field, and presents the goals, history and consequences of obtaining the genome sequence. With a foreword by James Watson, the book describes early achievements in DNA structure and human genome mapping, followed by a detailed account of the dramatic course of events that led to the sequencing the human genome ahead of schedule.

The style is very clear and the authors present not only the facts, but also the spirit of this dramatic race. Although it is probably too early for a full understanding, the authors have tried to assess the future impact of human genome sequencing on biology, medicine and society. The book contains very good illustrations and historical photographs of key participants of the project. The original article reporting the sequence obtained by the public consortium, with schematic gene maps and accompanying papers, makes up more than half of the book. It will be a pleasure for any scientist to have this milestone book in their personal library. ■

Daniel Cohen is at GENSET, 24 rue Royale, F75008 Paris, France.

Science in culture



Knowing neurons

A dynamic installation highlighting the fluidity of brain function

Martin Kemp

"The work is a homage to the beauty and sheer complexity of the structure and its steady development, its genetically-driven growth influenced by experience."

"The complexity and beauty of its structure... reflects the developmental constraints that shaped its growth."

These two quotes are from writers extolling the wonder of the neuron, as disclosed by modern neurobiology. The first is from Andrew Carnegie, one of the 'researching artists' featured in the exhibition "Head On. Art with the Brain in Mind" at the Science Museum in London (on view until 28 July 2002). The second is from Richard Wingate of the Medical Research Centre for Developmental Neurobiology, King's College London, in whose laboratory Carnegie was introduced first-hand to "the distributed circuitry whose logic remains one of the greatest mysteries of biology" (Wingate again).

Their shared experience was the result of one of eight such collaborations set up by the curators of the exhibition — Caterina Albano, Ken Arnold and Marina Wallace — under the aegis of the Wellcome Trust. The other artists are Osi Audu, Annie Cattrel, Catherine Dowson, Letizia Galli, Claude Heath, Gerhard Lang and Tim O'Riley.

In building up his three-dimensional visualization of the brain, Carnegie looked back to the pioneering work of Santiago Ramon y Cajal. Cajal used thin slices of tissue, impregnated with resin and stained by a method invented by Camillo Golgi, to construct an imaginative spatial picture of the cellular structure within the brain. Using Golgi stains, Cajal was able to infer the

composition of different brain regions and even to suggest the direction of information flow. Cajal and Golgi were awarded Nobel prizes in 1906.

The modern resources available to Carnegie, such as laser-scanning confocal microscopes and MRI scans, can be used to create compelling spatial arrays and to transcend the static pictures of morbid anatomy, capturing the three-dimensional dynamism of the living circuitry. Carnegie was drawn into the amazing worlds of the proliferation, migration and connectivity of neurons in the developing mind, particularly as drifting memories are laid down.

What Carnegie set out to capture in his installation, *Magic Forest* (above), was not an illustration of brain physiology but an evocation of the fluid flux that is the essence of neuronal transformation, as the growing cells extend their branches to communicate with companions near and far. Two projectors stand three metres apart on plinths at either end of a darkened space. They alternately project 160 slides on to three large gauze screens. As the images dissolve into one another, the forest grows and diminishes, comes and goes, builds and collapses, layer by layer, in an endless loop of generation and decay.

At these cutting edges of creative visualization, the tasks of the artist and the scientist both begin at the boundaries where knowledge runs thin. The artist gives vent to his awe through the magic of visual suggestion; the scientist through an insatiable urge to explain 'how'.

"It has been a breathtaking experience." The words are Carnegie's, but they could have been said by any scientist grappling with what Wingate calls the "slices, fragments and snapshots" that "remain, for the time being, the basis of our understanding of neuroanatomy".

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.

How plants fight dirty

Jack C. Schultz

Each of the world's 300,000 plant species is a target for attack from a range of nearly 400,000 species of plant-eating insects. Herbivory is among the Earth's most important interactions in terms of number of taxa, biomass and mass transfer, as well as evolutionary impact on plant traits, community structure and ecosystem function. The population biologists Paul Ehrlich and Peter Raven have even claimed that insect herbivory has generated much of terrestrial biodiversity. Yet we are now realizing that the interactions between plants and insects are far more dynamic than was previously thought, and involve much shared chemistry.

Insects are responsible for 15% of the world's crop losses; even in natural systems, they consume 10% of plant production each year. Herbivory is limited directly or indirectly by plant chemicals, which at first glance seem to have no role in normal plant metabolism but are highly active in animal tissues. Plant-herbivore interactions are often described as an ongoing biochemical warfare that occurs on an evolutionary timescale.

But our perspective of plant defences and plant-insect interactions is now changing. Plants are not static, chemically defended fortresses — they respond to attack with rapid, long-lasting, variable, and often specific biochemical, physiological and developmental changes. Plants respond differentially to many stimuli, including various insect species. Every class of constitutive chemical defence responds to a physical insect attack, or to insect regurgitant or saliva, over periods of minutes to days. The few 'stealthy' insect species that fail to elicit any response at all are now considered exceptional.

Several signalling pathways coordinate these responses. Fatty-acid signals (for example, oxylipins, which are synthesized

from linolenic acid released from membranes by lipases) regulate expression of defence-related genes and are central to most wound-mediated plant responses. Peptides, phenolics, terpenoids and classical plant hormones (such as cytokinins and ethylene) also can help to coordinate plant responses.

As plants recognize pathogens by detecting specific molecules, it surely follows that feeding insects must likewise present identifying chemistry to plant receptors. But few such response elicitors have been isolated from insects. Salivary digestive enzymes have some activity, as do several fatty-acid derivatives. The most complete study so far indicates that phospholipid/amino acid conjugates found in insect guts — and hence presumably in 'spit' (or regurgitant) — elicit what may be herbivore-specific volatile emissions from the host plant. These emissions in turn attract parasitoids, which kill the insect pests.

Plant-response elicitors, and the fatty-acid-based oxylipins used by plants themselves to organize defence, are similar to signals used in animal defence systems. Eicosanoids in animals (including insects) match plant oxylipins in terms of biosynthesis, structure, function and oxidative modification. These fatty-acid signals, including prostaglandins, leukotrienes and thromboxanes, mediate many immune and inflammation responses in animals. Eicosanoid receptors and their genes, as well as their structure-function relationships, have been characterized in vertebrates. Much of their activity is attributed to structural features that are shared by plant oxylipins.

Other chemical signals that are shared across the plant and animal kingdoms include peptides and glycoproteins, nitric oxide, reactive oxygen species, neurotransmitters and plant growth hormones (auxins and cytokinins). Plant phenolics and steroids interact with animal steroid receptors, and many plants produce molecules that act as insect hormones. The signalling activities of many of these molecules are known only in one of the kingdoms, but this probably reflects an absence of evidence — or study — rather than evidence of absence.

Plant defence responses require differential gene expression in signalling, sensory and metabolic pathways. It now seems that 800–1,500 differentially regulated genes are involved in a plant's response to a single insect, a complexity that matches that of animal responses.

Searching plant and animal genomes for sequences that are likely to share functions has the potential to revise our view of plant defences completely — for example, a sequence that encodes a putative glutamate

Biochemical ecology

Plant-herbivore interactions are frequently described as an ongoing biochemical warfare that occurs on an evolutionary timescale.

receptor has now been found in the genome of the mustard weed (*Arabidopsis thaliana*). Ecologists have long known that plants produce neurotransmitters, including glutamate, that are presumably involved in defence against animals. The discovery of such receptors in plants suggests that such 'secondary metabolites' are more than purely defensive, and that plants may be able to sense and respond to common animal signals.

There are important ecological and evolutionary implications of plants and animals sharing signalling systems. Theoretically, each could manipulate the other's detection and response mechanisms. Insects present familiar signals to plants when they attack, and plant food contains chemicals that interfere with signalling in animals. This latter observation is the basis of phytopharmacology, but is unappreciated in ecological and evolutionary contexts. Although the phylogenetic distance between plants and herbivores is great, the organisms share a common code-book, which explains how plants might identify insects, how both plants and insects might 'jam' each other's signalling, and how plants could wreak havoc in insects by interfering with immune responses, reproduction and behaviour.

This new view of herbivory, emphasizing rapid, dynamic behaviours, shifts our focus from the chemical composition of plant tissues to the factors that elicit change and the signals that coordinate it. Here there are as many similarities between plants and insects as there are differences, raising the possibility of much 'signal stealing' and 'biochemical espionage' between their two kingdoms. Perhaps, in terms of perception, signalling and biochemical behaviour, plants are really just very slow animals after all.

Jack C. Schultz is in the Department of Entomology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

FURTHER READING

- Ehrlich, P. R. & Raven, P. H. *Evolution* **18**, 586–608 (1964).
- Karban, R. & Baldwin, I. T. *Induced Responses to Herbivory* (Univ. Chicago Press, Chicago, 1997).
- Shiu, S.-H. & Bleecker, A. B. *Proc. Natl Acad. Sci. USA* **98**, 10763–10768 (2001).
- Chiu, J., DeSalle, R., Lam, H. M., Meisel, L. & Coruzzi, G. *Mol. Biol. Evol.* **16**, 826–838 (1999).



A frog hopper nymph coated in protective froth elicits a hugely complex defence by its plant host.

PAPILIO/CORBIS

The north–south martian divide

Peter Gierasch

Some of the differences between the northern and southern hemispheres of Mars may stem from asymmetry in the planet's atmospheric circulation, and the resulting distribution of water and dust.

Mars is an asymmetrical planet. Its northern hemisphere is relatively smooth and free of craters, with a large, permanent ice-cap that is mainly composed of water. The southern polar cap is smaller and contains carbon dioxide, and the surrounding terrain is heavily cratered (Fig. 1). Hemispheric dichotomies are not unusual in the Solar System. The Moon, for example, has extensive lava plains on the side facing the Earth and is heavily cratered on the other side. Unlike the Moon, however, Mars has an atmosphere that can transport dust and water vapour around the planet. So one might expect that, over the long term, mixing would lead to a balanced distribution of polar ice. But this is not what we see.

On page 298 of this issue¹, Richardson and Wilson propose that atmospheric transport of water and dust, rather than tending to redistribute material uniformly, in fact acts as a pump across the martian equator and contributes to the north–south asymmetry in ice coverage. It may also produce an imbalance in the distribution of surface dust deposits, although the absence of depth measurements of the deposits means that this possibility remains speculative. Mars has no ocean, and therefore the atmosphere is the only medium for transportation on the planet. So Richardson and Wilson's suggestion is important.

Mars has an elliptical orbit around the Sun; its eccentricity is 0.093 (0 indicates a perfectly circular orbit). The axis of rotation is tilted away from the normal by 25.1°. So Mars has seasons similar to the Earth's, but they are exaggerated and more asymmetrical. At present, summer in the martian southern hemisphere occurs near perihelion — that part of its orbit when the planet is closest to the Sun. At this point, Mars is about 20% closer to the Sun than it is during the northern hemisphere summer.

Thus, one possible cause of the asymmetry is that southern hemisphere summers are much warmer than northern hemisphere summers. Another possibility lies in the fact that the terrain in the southern hemisphere is on average higher than that in the north. But study of these factors has produced no widely accepted explanation for the asymmetries in water distribution, and the observations remain puzzling^{2–4}. For example, a thermal explanation for the paucity of water ice in the south might at first appear promising

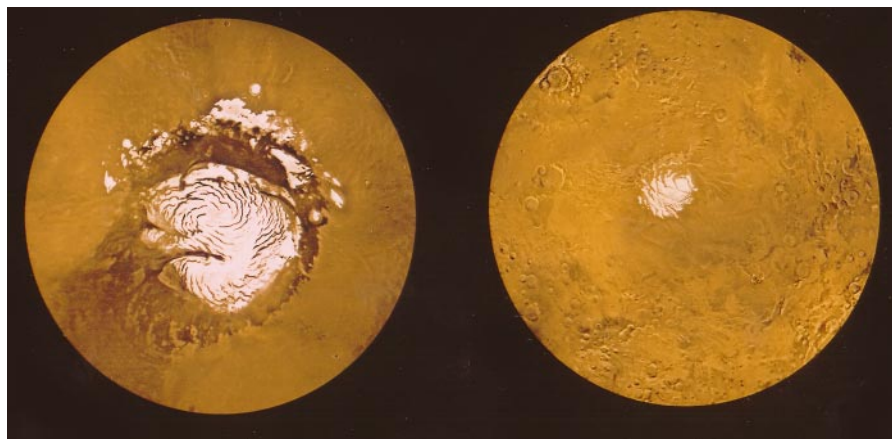


Figure 1 Mars, north and south. The extensive northern polar cap (left) is composed of water, and the surrounding terrain is relatively smooth. The southern cap is smaller, and is thought to contain both water and carbon dioxide. Cratering in the southern hemisphere is comparatively heavy. These pictures are projections of images taken by the NASA Viking Orbiter spacecraft, and show each hemisphere from 65° latitude to the poles.

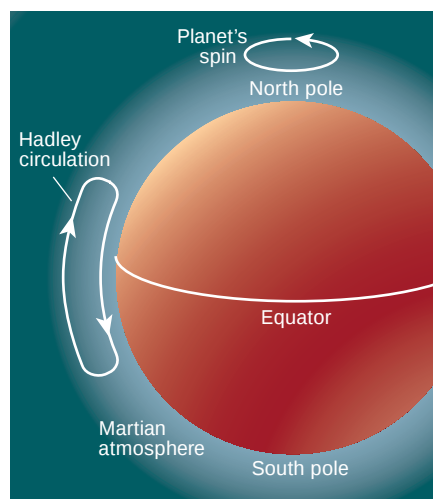


Figure 2 A cross-equator Hadley circulation during summer in the southern hemisphere on Mars. The flow changes direction during summer in the northern hemisphere, but Richardson and Wilson¹ propose that because the terrain in the planet's southern hemisphere tends to be higher than that of the northern hemisphere, the southern summer circulation dominates. The asymmetric flow, on an annual average, is northward at high altitudes in the atmosphere, and southward lower down — perhaps explaining certain inter-hemispheric differences on Mars.

because of the warm southern summers. But this factor is at least partly offset by the shorter duration of the southern summer as a result of Mars's rapid orbital motion near perihelion. Furthermore, subtle factors, such as the reflectivity of dusty ice to sunlight, can also have strong effects^{2,3}.

Richardson and Wilson¹ present a new idea to explain how the atmosphere might cause the north–south asymmetry on Mars. Using a set of numerical circulation experiments, they show that, on average during the year, the atmosphere circulates across the equator with northward flow at high levels in the atmosphere and southward flow at low

levels. This circulation — known as a Hadley circulation (Fig. 2) — is produced by heating at the latitude on a planet where the noon Sun is directly overhead.

The annual average Hadley circulation on the Earth was originally viewed as a major component of the general circulation of the atmosphere, with air rising at the Equator and sinking towards the poles. More recently it has been realized that the strongest Hadley circulations occur at the solstices, when the Sun is at a maximum distance from the Equator⁵. At mid- and high latitudes, Hadley circulations are strongly inhibited by Coriolis forces, which result from the Earth's

rotation. These forces turn the flow at right angles to its original path, rather than allowing it to move directly down the pressure gradient. Coriolis forces vanish at the Equator. So the strongest Hadley circulation occurs across the Equator at solstices, when the gradient of solar heating between the two hemispheres is largest.

On Mars, the Hadley circulation at the solstices is stronger than that on the Earth and spans a much wider range of latitude, reaching beyond 50° N at the southern hemisphere summer solstice⁶. The mass of the martian atmosphere (per unit area) is only about 1.3% that of the Earth's, and is composed primarily of carbon dioxide, a strong absorber of infrared radiation. Consequently, the martian atmosphere responds strongly to temperature gradients created by solar radiation.

In the new simulations¹, the cause of the north-south asymmetry in the Hadley circulation is the difference in the height of the terrain in the two martian hemispheres. This elevation gradient across the planet's equator remains fixed, of course, whereas the parameters relating to its orbit and spin vary and perhaps cancel out hemispheric differences over time. The origin of the difference in elevation, which averages about 5 km (ref. 7), is unknown. One hypothesis is that, early in martian history, global-scale convective movement of the planet's mantle — the layer beneath the surface layer, or crust — pulled crustal material towards the south, thickening the crust there⁸. Another is that a large impact removed crustal material from the northern hemisphere⁹.

Richardson and Wilson suggest two mechanisms that might be driving the atmospheric flow, but do not claim to have fully analysed its origin. One idea is that the warm surface, which is at relatively high elevations in the southern hemisphere, heats the atmosphere up at elevations where it would be cool if the surface were not elevated. Since the surface in the northern hemisphere is not elevated, this leads to a temperature gradient from north to south. Another possible explanation is that the upward slope of the warm surface from north to south across the martian equator favours rising atmospheric movement during the southern solstice over sinking movement during the northern solstice. But whatever the detailed mechanism, Richardson and Wilson's numerical experiments, which they performed with and without incorporating the difference in terrain height between the two hemispheres, convincingly show that this difference is the cause of the annual mean circulation.

Trace constituents of the atmosphere, such as water, dust and possibly certain gases, generally occur in a concentration gradient with height. So a cross-equator Hadley circulation can have a strong effect on transport between the hemispheres by acting as a pump. In the case of water, Richardson and

Wilson argue that its concentration will generally be highest in the warm summer hemisphere, where its vapour pressure is highest, and therefore that it will be most abundant in the upper part of the cross-equator circulation, flowing from summer towards winter hemispheres. So this bias could stabilize the water ice-cap in the northern hemisphere.

Understanding climate dynamics is a complicated and tricky business, especially on a planet where the buried geological record of climatic effects is beyond our reach. Nonetheless, Richardson and Wilson have pointed out a new constraint on the interplay between the surface of Mars and its atmosphere, one that should help us to understand the planet much better. ■

Peter Gierasch is in the Department of Astronomy, Cornell University, 318 Space Sciences Building, Ithaca, New York 14853, USA.

e-mail: gierasch@astro.cornell.edu

1. Richardson, M. I. & Wilson, R. J. *Nature* **416**, 298–301 (2002).
2. Kieffer, H. H. & Zent, A. P. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 1180–1218 (Univ. Arizona Press, Tucson, 1992).
3. Jakosky, B. M. & Haberle, R. M. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 969–1016 (Univ. Arizona Press, Tucson, 1992).
4. James, P. B., Kieffer, H. H. & Paige, D. A. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 934–968 (Univ. Arizona Press, Tucson, 1992).
5. Lindzen, R. S. & Hou, A. Y. *J. Atmos. Sci.* **45**, 2416–2427 (1988).
6. Smith, M. D., Pearl, J. C., Conrath, B. J. & Christensen, P. R. *J. Geophys. Res.* **106**, 23929–23945 (2001).
7. Smith, D. E. *et al. Science* **284**, 1495–1503 (1999).
8. Wise, D. U. *et al. Icarus* **38**, 456–472 (1979).
9. Wilhelms, D. E. & Squyres, S. W. *Nature* **309**, 138–140 (1984).

Neurobiology

Ready to unlearn

Shigeru Kitazawa

After rabbits learn to associate a tone with a small shock near the eye, they blink when they hear the tone. Learning requires activation of nerve fibres known as climbing fibres. Inhibition of these fibres leads to 'unlearning'.

Many of us were taught about Pavlov's classic experiments in school: when a dog is repeatedly fed after a neutral tone is sounded, the animal soon learns to salivate when it hears the tone. There are, of course, other pavlovian responses, and one that is commonly used in laboratory studies is the 'eye-blink' response. A brief shock near the eye causes 'innate', defensive blinking. When a brief tone is repeatedly paired with the shock, the tone comes to elicit a 'conditioned' blinking response that begins just before and peaks at around the time that the shock would be expected.

Just as important as learning an association like this, however, is forgetting it when necessary. So, if the tone is sounded without a shock, the rate and amplitude of blinking gradually decrease until the tone no longer evokes a response. This process is known as extinction and, on page 330 of this issue, Medina *et al.*¹ reveal how it comes about.

Several brain regions, which are either within or connected to the cerebellum, are required to learn the conditioned eye-blink response^{2–4} (Fig. 1). A shock elicits defensive blinking by setting off electrical impulses through simple neuronal circuits (known as reflex arcs) in the brain stem. By contrast, a tone causes conditioned blinking by sequentially activating the pons, mossy fibres and interpositus nucleus, among other parts of the brain.

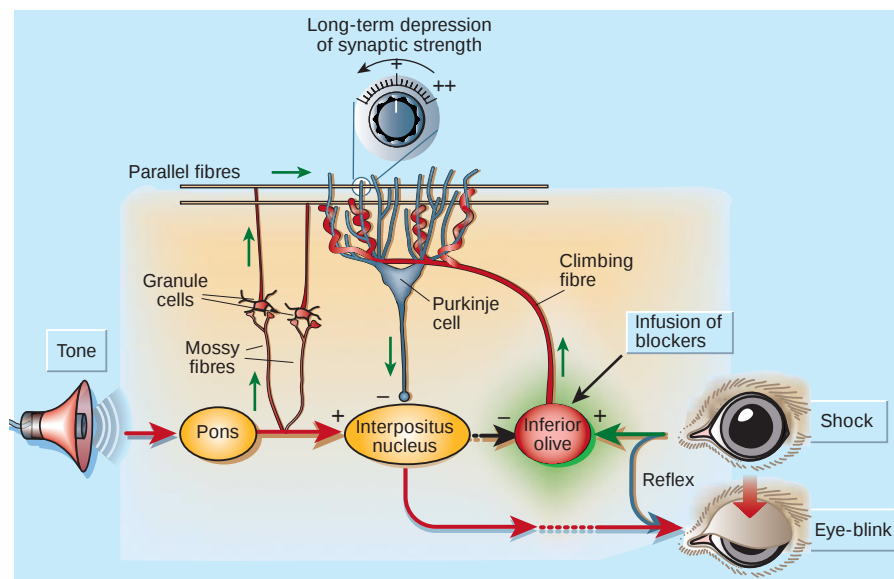
Before the conditioned response has been 'acquired', impulses set off by the tone are blocked at the interpositus nucleus by so-called Purkinje cells. When the tone and shock are combined, electrical impulses are

conveyed in parallel to the Purkinje cells, the tone by way of mossy fibres, granule cells and parallel fibres, and the shock via the inferior olive and climbing fibres. The joint activation of parallel fibres and climbing fibres induces long-term depression — which is thought to involve long-lasting molecular changes — of the connections (synapses) between the parallel fibres and Purkinje cells⁴. This reduces the activity of the Purkinje cells. So, when the tone and shock are repeatedly paired, the activity of the Purkinje cells is progressively reduced, and so too is the inhibition of the interpositus nucleus by the Purkinje cells. This means that the tone can now act via the interpositus nucleus to evoke conditioned blinking.

Medina *et al.*¹ now reveal that extinction, which is no less important than acquisition, is not a passive process of 'forgetting', but rather an active 'unlearning' process that is driven by inhibition of climbing fibres. After training rabbits to blink in response to a tone, the authors infused various different fluids into the inferior olive. As a control they used artificial cerebrospinal fluid; the conditioned response gradually disappeared in 60 to 70 'tone-alone' trials — as it did when no fluids were infused. But when Medina *et al.* used picrotoxin, which blocks the inhibitory neurotransmitter GABA, the rabbits continued to blink even after 100 tone-alone trials. So blocking inhibitory inputs to the inferior olive prevents extinction. In contrast, blocking excitatory inputs to the inferior olive induced extinction even in the presence of both tone and shock¹. These results suggest that inhibition of the inferior olive, and

The similarity between malignant disease and tuberculosis has led numerous investigators to seek for an organism which would bear the same causative relation to cancer as the tubercle bacillus does to tuberculosis... The main point of difference between the adherents of the parasitic theory of the origin of cancer and their opponents centres upon the significance of certain undoubted microscopic appearances, chiefly of the growing portions, of cancerous growths. Some observers maintain that these microscopical appearances represent an organism of a protozoic type, others regard them as due to degeneration of the cancer cells. The majority, however, of microscopists do not regard the presence of a parasite in cancerous growths as proved. In the case of sarcomata, the parasite is supposed to be, not of animal, but of vegetable origin... If we turn from the study of the hypothetical cancer parasite to a consideration of the influence of general climatic conditions upon the incidence of cancer, we shall be treading upon more certain ground. The existence of so-called "cancer houses" seems to rest upon very strong evidence.

Sir Lawrence Bragg delivered the thirty-sixth (1952) Guthrie Lecture of the Physical Society on March 12, when he discussed "X-ray Analysis of Proteins". Sir Lawrence said that the attempt to discover the atomic arrangement in the protein molecule seems very ambitious. Ever since Bernal first showed that crystals of protein give X-ray diffraction pictures, it has been clear that a protein molecule of a given type is a structure with a definite individual form; the X-ray diffraction spots are very sharp and reproducible and extend to regions corresponding to interatomic distances. The molecules are, however, of great complexity. It has been a triumph of X-ray analysis to pass from simple substances like rock-salt to such molecules as strychnine or penicillin with about one hundred atoms. Attempts are now being made to analyse a molecule such as haemoglobin, which contains fifteen thousand atoms. The reward, if an analysis were completed, would be great, because the determination of any one protein would undoubtedly cast a flood of light on the character of these bodies, which Nature has selected as the basis of living matter. From *Nature* 22 March 1952.



What are the molecular mechanisms that underlie extinction? Long-term potentiation^{4,7} (strengthening) of synapses between parallel fibres and Purkinje cells is one possibility. But this would not 'undo' the long-term depression at these synapses at the molecular level, because long-term potentiation

Medina *et al.*'s work¹ also has broader implications. Climbing fibres are thought to be used more generally when learning movements, to convey to the cerebellum that errors are being made⁴. In the case of conditioned blinking, the signal for extinction (the inhibition of climbing fibres) might be interpreted as telling the cerebellum about a movement error — that is, the eye is blinking unnecessarily. One benefit of using the output from the interpositus nucleus to generate the extinction signal would be to save time in calculating this movement error. The error signal would begin before the eye actually blinks, if neuronal projections from the interpositus nucleus do indeed inhibit the inferior olive^{4,5,8,9}. If the outputs from other cerebellar nuclei are used in the same way in calculating movement errors, then errors might be predicted before they are made, as

observed in monkeys reaching towards a target¹⁰.

Finally, it is now clear that a decrease in climbing-fibre activity is the key to erasing conditioned blinking, whereas an increase in this activity underlies the acquisition of this response. Such increases and decreases might be interpreted as representing an unexpected shock (or a 'negative reward'), and an unexpected relief from that shock (a positive reward) — that is, a reward-prediction error¹¹. In this context, acquisition of this conditioned response in the cerebellum could be regarded as a typical example of reinforcement learning¹¹, where a discrepancy between the predicted and actual reward is needed to drive learning. So the strategies adopted by Medina *et al.*¹ could be directly applied to studying whether increases and decreases in the activity of dopamine-releasing neurons, which are thought to represent

reward-prediction errors in many situations¹¹, are the keys to adopting or abandoning behaviours more generally. ■

Shigeru Kitazawa is at the Neuroscience Research Institute, National Institute for Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba 305-8568, Japan.

e-mail: kitazawa-s@aist.go.jp

1. Medina, J. F., Norez, W. L. & Mauk, M. D. *Nature* 416, 330–333 (2002).
2. Kim, J. J. & Thompson, R. F. *Trends Neurosci.* 20, 177–181 (1997).
3. Mauk, M. D., Garcia, K. S., Medina, J. F. & Steele, P. M. *Neuron* 20, 359–362 (1998).
4. Ito, M. *Physiol. Rev.* 81, 1143–1195 (2001).
5. Kim, J. J., Krupa, D. J. & Thompson, R. F. *Science* 279, 570–573 (1998).
6. Fanselow, M. S. *Neuron* 20, 625–627 (1998).
7. Sakurai, M. J. *Physiol. Lond.* 394, 463–480 (1987).
8. Hansel, C., Linden, D. J. & D'Angelo, E. *Nature Neurosci.* 4, 467–475 (2001).
9. de Zeeuw, C. I. *et al. Trends Neurosci.* 21, 391–400 (1998).
10. Kitazawa, S., Kimura, T. & Yin, P. B. *Nature* 392, 494–497 (1998).
11. Schultz, W. *Nature Rev. Neurosci.* 1, 199–207 (2000).

Molecular physiology

Protecting the heart

Mark T. Nelson and Gerald M. Herrera

Abnormal enlargement of the heart muscle can be life-threatening. Unexpected insight into the underlying molecular mechanisms has come from investigating a gene that regulates cardiac calcium levels.

The rhythmic contraction and relaxation of the heart depends on the dependable rise and fall of the level of calcium ions inside heart muscle cells. Although the basic mechanisms and much of the cellular machinery that produces these changes have long been known, the roles of several regulatory proteins are poorly understood. It is often difficult to work out what a protein does, especially if it serves in several pathways and shows cell-type-specific properties. But, by manipulating an organism so that it lacks the gene encoding the protein of interest, and then assessing the effects, a researcher can usually learn a great deal. Often, to some extent, the effects can be predicted. But sometimes the results are unexpected, and so it was with Xin and colleagues' study¹ of the FKBP12.6 protein. The authors were aiming to clarify how this protein functions in the heart. As they report on page 334 of this issue, they found that male mice lacking this protein have abnormally enlarged hearts but females do not. This gender-specific protection of the heart seems to stem from the presence of oestrogen.

What made FKBP12.6 a candidate for an involvement in heart function? First, disruption of a closely related protein, FKBP12, has profound effects on heart development in mice². Second, both of these proteins are thought to be involved in calcium-triggered signalling pathways, which have several inter-related roles in the cardiovascular system. For

example, calcium regulates the contraction of heart muscle cells and of smooth muscle cells in blood vessels. It also controls the release of nitric oxide and other factors from vascular endothelial cells, the cells that line blood vessels; nitric oxide in turn has the effect of dilating blood vessels, so lowering blood pressure. (High blood pressure is, of course, potentially life-threatening: it can lead to strokes, heart attacks and other forms of heart failure.) And finally, calcium is involved in activating certain gene-transcription factors, such as NFAT, that have been implicated in heart failure³ and vascular function^{4,5}. For instance, abnormally high cellular calcium levels lead to the activation of NFAT, which in turn switches on a programme of gene expression that may lead to heart failure³.

Many calcium-triggered signalling pathways work through channels known as ryanodine receptors, which sit in the outer membrane of an intracellular calcium store, the sarcoplasmic reticulum⁶. In response to an influx of extracellular calcium ions into the cell, these channels can open to release calcium from the sarcoplasmic reticulum into the body of the cell, leading to a further rise in the intracellular calcium concentration that triggers muscle contraction. In other words, the ryanodine receptors link stimulation of the cell to muscle contraction; this is known as excitation–contraction coupling.

Of course, maintaining rhythmicity then

requires that contraction give way to relaxation. To stop the heart from contracting, ryanodine receptors must shut to prevent the intracellular calcium levels rising any further. This is where the FKBP proteins are thought to be involved. As mentioned above, disruption of the FKBP12 gene leads to alterations in heart structure and function in mice². It also increases the probability that cardiac and skeletal-muscle ryanodine receptors will open *in vitro*. In other words, FKBP12 seems to regulate the flow of calcium through ryanodine receptors, and is therefore involved in excitation–contraction coupling.

But what about FKBP12.6? This protein is thought to associate selectively with the cardiac ryanodine receptor, so is it involved in excitation–contraction coupling in heart muscle? To find out, Xin *et al.*¹ knocked out the encoding gene in mice. They discovered that ryanodine receptors in isolated cardiac muscle cells — from both male and female mice — became more active, as measured by an increase in the duration of calcium 'sparks' and in the release of calcium from the sarcoplasmic reticulum. Thus, FKBP12.6 does seem to regulate cardiac ryanodine receptors.

These effects on isolated heart cells are perhaps not surprising. But an unexpected finding was revealed when the authors examined intact hearts from FKBP12.6-deficient mice. Unlike FKBP12-deficient mice, which die either as embryos or soon after birth², the FKBP12.6-lacking mice survive and mature¹. But Xin *et al.* detected a marked and unanticipated gender difference: the male, but not female, mice lacking FKBP12.6 developed cardiac hypertrophy — a potentially pathological increase in cardiac mass — reminiscent of the more severe effects of disrupting FKBP12.

These intriguing results suggest that the female FKBP12.6-lacking mice are somehow protected from cardiac hypertrophy. Xin *et al.* decided to test whether oestrogen — a predominantly female hormone — lies at the root of this protection. Oestrogen was already known to have a 'cardioprotective' effect⁷, in part because it increases the release of nitric oxide from vascular endothelial cells^{7–9}, leading to blood-vessel dilation. Moreover, if the gene encoding oestrogen receptor- β is disrupted in mice, high blood pressure ensues¹⁰. So Xin *et al.*¹ treated female FKBP12.6-deficient mice with the breast-cancer drug tamoxifen, which can antagonize oestrogen receptors to block the effects of circulating oestrogen. These females developed comparable cardiac hypertrophy to the male FKBP12.6-lacking animals, suggesting that oestrogen usually protects female FKBP12.6-deficient mice from cardiac hypertrophy.

How might this protection work? As mentioned, oestrogen increases blood-vessel dilation, lowering blood pressure. It is plausible that FKBP12.6 does the same —

albeit in a different way, by modulating ryanodine receptors in blood vessels. If so, the absence of FKBP12.6 would lead to an increase in blood pressure, requiring the heart to pump more forcefully. And the predicted outcome of more forceful cardiac contraction in the face of lasting high blood pressure is enlargement of the heart. Indeed, Xin *et al.* found that male, but not female, FKBP12.6-null mice had high blood pressure¹. This is consistent with the idea that cardiac hypertrophy in the males is a secondary consequence of elevated blood pressure, and that oestrogen opposes the effects of a lack of FKBP12.6 on blood pressure.

There are other possibilities, however. The increases in intracellular calcium levels seen in FKBP12.6-deficient animals might lead directly to an increase in the activity of the calcium-dependent transcription factor NFAT. Alternatively, because FKBP12 and FKBP12.6 not only regulate calcium release through ryanodine receptors but are also important components of the calcineurin protein complex, which regulates NFAT activity, it is possible that disruption of FKBP12 or FKBP12.6 in the heart or blood vessels leads to the activation of NFAT

through effects on calcineurin. The activation of NFAT by either of these two mechanisms might induce a gene programme that causes an increase in blood pressure or cardiac hypertrophy. It is conceivable that oestrogen decreases the expression or activity of NFAT or its target genes in the blood vessels or hearts of females. Whatever the answer, an understanding of the molecular mechanisms at work will no doubt be crucial to comprehending gender-based differences in cardiovascular function. ■

Mark T. Nelson and Gerald M. Herrera are in the Department of Pharmacology, University of Vermont, Given Medical Building, Burlington, Vermont 05405-0068, USA.

e-mail: mtnelson@uvm.edu

1. Xin, H.-B. *et al.* *Nature* 416, 334–337 (2002).
2. Shou, W. *et al.* *Nature* 391, 489–492 (1998).
3. Molken, J. D. *et al.* *Cell* 93, 215–228 (1998).
4. Graef, I. A., Chen, F., Chen, L., Kuo, A. & Crabtree, G. R. *Cell* 105, 863–875 (2001).
5. Stevenson, A. S., Gomez, M. F., Hill-Eubanks, D. C. & Nelson, M. T. *J. Biol. Chem.* 276, 15018–15024 (2001).
6. Marks, A. R. *Physiol. Rev.* 76, 631–649 (1996).
7. Mendelsohn, M. E. & Karas, R. H. *N. Engl. J. Med.* 340, 1801–1811 (1999).
8. Wellman, G. C., Bonev, A. D., Nelson, M. T. & Brayden, J. E. *Circ. Res.* 79, 1024–1030 (1996).
9. Geary, G. G. *et al.* *J. Appl. Physiol.* 91, 2391–2399 (2001).
10. Zhu, Y. *et al.* *Science* 295, 505–508 (2002).

Materials science

Breaking the neural code

Adam Curtis

The precise information that is conveyed between nerve cells remains unknown. Networks of nerve cells grown on silicon chips, using a polyester as a guide, may bring us closer to translating the elusive neural language.

Within a nerve cell, or neuron, differences in electrical potential encode information, and messages can be passed on to other neurons through connections that may be chemical or electrical. But, like nineteenth-century linguists contemplating the hieroglyphs of the Rosetta Stone, we have only vague clues as to the symbolism and language, let alone the grammar or syntax, of neural messages. We will never dig up a neurobiological Rosetta Stone—but we might be able to synthesize our own. In a notable paper in *Advanced Materials*, Merz and Fromherz¹ have taken us at least one stage towards this goal, and their work brings with it the promise of further discoveries.

To decipher the neural code, the aim is to grow networks of neurons in culture, with each neuron connected to others through synapses that form at the ends of the neuron's tendril-like extensions, or neurites. We need to reproduce the connectivity that is seen anatomically in nervous systems and which, when destroyed, leads to loss of correct function. We suspect that the message passing across a synapse carries instructions for how the information should be processed and

possibly also routed, and in many cases we know what the correct (or at least normal) input and output are. What we would like to test are the effects of different connectivities and input messages on the neuronal output.

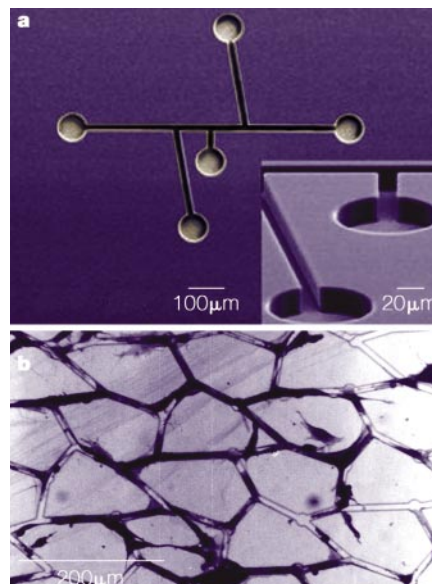


Figure 1 Artificially grown networks of neurons could be the key to deciphering the intricate code of neural messages. a, Merz and Fromherz¹ etched pits and grooves in a polyester–silicon material to control the growth of neurites into a functioning network; pits 70–80 μm in diameter match the size of the pond-snail neurons used in the experiment, and grooves up to 15 μm wide and 30 μm deep confine and guide the neurites into forming synaptic connections between neurons. b, Following a pattern suggested by Wilkinson⁶, Sandison *et al.*⁵ grew this neuronal network in grooves in silica—although the difficulty of confining the growth within the structure is apparent.

a, M. MERZ & P. FROMHERZ / b, A. CURTIS

glass microelectrodes, although their earlier work had used extracellular electrodes, based on field-effect transistors to amplify the weak electrical signals. Extracellular electrodes have an advantage over the intracellular kind because trapping neurons with fixed electrodes, or piercing them with a micro-electrode, can kill them. In fact, my own group has recently invented a manipulable extracellular electrode array⁵ that can be rapidly moved from site to site. The electrode patterns can be matched to that of the network under test, although the array's extracellular position does reduce the magnitude of the signal that can be extracted.

The growth technique of Merz and Fromherz, when combined with advances in electrode design, should open the way to investigating how connectivity and patterns of connection modify neural processing. In other words, it should be possible to test the consequences of specific changes in connectivity or in the electrically injected message on the output. Furthermore, as Merz and Fromherz¹ point out, their system should allow the role of subthreshold signals to be assessed — these are events that do not lead directly to transmission of a neural message, but which may influence the electrical potentials of nerve cells.

With the tools in place, the outstanding question is how a neuronal network could be designed. There are two sensible approaches: first, to try to mimic what is already found in nature; and second, to formally develop logically designed networks. Judith Wilkinson (quoted in ref. 6) has applied graph and tiling theory to design networks that should be able to propagate and compare messages (Fig. 1b). These are probably the only *in vitro* systems that feature branching networks that are just like those in real nervous systems. Such logically designed networks are just now being realized.

Despite the progress reported by workers such as Merz and Fromherz, we should still proceed cautiously — for instance, we do not even know whether there are (or aren't) several types of neural coding, or language, in a single animal. But, as in other code-breaking efforts, there will probably be a long period of data collection — revealing more and more of the neurobiological Rosetta Stone — until we have enough information to start interpreting the neural code. ■

Adam Curtis is at the Centre for Cell Engineering, IBLS, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, UK.

e-mail: a.curtis@bio.gla.ac.uk

1. Merz, M. & Fromherz, P. *Adv. Mater.* **14**, 141–144 (2002).
2. Lauer, L., Klein, C. & Offenhausser, A. *Biomaterials* **22**, 1925–1932 (2001).
3. Maher, M. P. *et al. Cell. Eng.* **37**, 110–118 (1999).
4. Syed, N. I., Bulloch, A. G. M. & Lukowiak, K. *Science* **250**, 282–285 (1990).
5. Sandison, M., Curtis, A. S. G. & Wilkinson, C. D. W. *J. Neurosci. Methods* **114**, 63–71 (2002).
6. Wilkinson, C. D. W. & Curtis, A. S. G. *Physics World* **12**, 45–48 (1999).

Immunology

The Wright stuff

Walter Gratzer

White cells in the bloodstream seek out and destroy invading bacteria. The explanation for the actual killing mechanism turns out to be wonderfully more subtle than previously thought.

On page 291 of this issue, Segal and colleagues (Reeves *et al.*)¹ cast new and unexpected light on one of the cornerstones of the immune system. Their observations overturn the established wisdom that has guided the study of phagocytosis and pose questions about the role of reactive oxygen species in destroying invading pathogens.

It is 120 years since Elie Metchnikoff discovered phagocytes — cells that engulf and devour bacteria and other invaders of the bloodstream. This 'innate immunological response' is the body's first line of defence against such violations, and it depends predominantly on the most abundant class of phagocytic cells, the most numerous white cells in the blood, the neutrophils. Sufferers from the rare, hereditary chronic granulomatous disease (CGD), whose neutrophils

are defective, die from bacterial and fungal infections if they are not treated. (Metchnikoff later came to believe that, with the passage of the years, phagocytic cells become incontinent, rampage through our tissues, gnaw at our vitals and cause the catastrophe that we recognize as ageing.)

Neutrophils surge towards the site of an infection and engulf bacteria by a well-characterized process (opsonization) that requires blood serum proteins. They then destroy the involuntary guest trapped within a phagocytic vacuole. The cytoplasm of the neutrophil is rich in granules (whence its original description as a granulocyte), which, on activation, discharge their contents inside the phagocytic vacuole.

The picture of the killing process that has dominated thinking over the past 40 years is that the lethal agents are the highly reactive

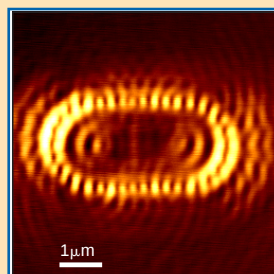
Quantum optics

Light corralled

Traditional optical experiments, such as splitting rays of light into various colours with a prism, have had the attraction of being visible to the naked eye. Modern methods of confining light within microscopic structures, and tailoring its interaction with matter on atomic scales, are taking optics into the quantum realm. Making the results visible is not straightforward. But a beautiful example comes from C. Chicanne *et al.*

(*Physical Review Letters* **88**, 097402; 2002), who have designed a way of taking snapshots of intricate light interference patterns in tiny photonic structures. One of their snapshots is shown here.

The images are reminiscent of the 'quantum corrals' for electrons, first created by Donald Eigler and colleagues in 1993, and produced with the scanning tunnelling microscope (STM). This instrument probes surfaces with a sharp needle,



which can also be used to move loose atoms around on a metallic surface and position them into a closed loop. The electrons confined within these corrals interfere with one another and produce beautiful patterns, which the STM images make manifest. These experiments have been highly instructive for illustrating the quantum-mechanical principle that electrons can behave as waves.

Chicanne *et al.* now present an optical analogue of the quantum corral. Light has wave properties, of course, but interpreting the interference effects is not straightforward

within a radius around a light source that is comparable to the wavelength of the light itself. To study this zone, Chicanne *et al.* made use of a relative of the STM, the scanning near-field optical microscope. Here, the surface probe is an optical fibre tapered to a sharp end that also illuminates the sample.

Much as in the experiments by Eigler and colleagues, Chicanne *et al.* first created corrals by carefully positioning particles of gold in a loop, in this example in the shape of a stadium. The optical corral is then imaged by scanning the fibre over the surface while collecting light transmitted through the transparent sample. The result is images of light interference patterns in a microscopic structure. The principles involved will help direct future observations, and tailored distribution, of light at atomic dimensions.

Liesbeth Venema

Daedalus

Fast forgetting

Botulinum toxin, the deadly nerve poison that inhibits the neurotransmitter acetylcholine, has many cosmetic uses these days. In extremely small quantities, it cancels facial tics, erases forehead worry-lines, and so on. The effect lasts for weeks or months, until the toxin decays; repeated treatments can be permanent. Daedalus sees psychiatric implications. The brain is the repository of many inappropriate or downright harmful reflexes, memories and sensory illusions. They could well be eliminated, if only we knew where they were stored.

Acupuncture and reflexology both claim that the whole body is mapped onto the skin surface. They work, if they do, by 'gating' nerves from the skin, launching signals that interfere with or override messages from the organs being treated. Nerves are tubes. Once inside a nerve, a virus or small particle is safe from immunological surveillance. Rabies, inflicted by the bite of a rabid animal, injects a virus at the bite site that slowly ascends the local nerve and then multiplies in the brain. So Daedalus will inject botulinum toxin into an acupunctural or reflexological skin location, chosen under psychiatric examination, and hope that it travels to the site of the psychiatric trouble. Unlike a virus, the toxin will not multiply and spread. Thus, to eliminate a troublesome memory the patient will recall it in detail, and the therapist will find which area of skin has been sensitized. Indeed, even if the patient cannot recall the memory (it may have been repressed), a sensitized area of skin should act as an indicator.

The toxin will be injected into the nerve as microencapsulated particles. By the time the encapsulation decays, the toxin should be on site. With luck the particles will slowly ascend that nerve alone, and travel exactly to that part of the brain concerned with the chosen memory. The process may take some time; but many psychiatric drugs also take a long time to act.

Psychiatry should be made much more precise. Bad memories, complexes, illusions such as tinnitus, each could be addressed exactly and removed without interfering with anything else. Even better, the effect should decay in a month or two with the toxin. The therapist will thus learn what other mental effects depend on the eliminated symptom. Repeated treatments could then be targeted more exactly.

David Jones

superoxide ions and hydrogen peroxide, generated by enzymes in the vacuole². Not only were these held to act directly on the microbe, but the hydrogen peroxide also served as substrate for destructive halogenation reactions, catalysed by another enzyme, myeloperoxidase. This scheme received what appeared to be conclusive support from the discovery of CGD patients, whose neutrophils engulfed bacteria but failed to unloose the toxic agents. Much later, moreover, it was reported that genetically modified mice lacking myeloperoxidase are highly sensitive to infection by the common fungal pathogen *Candida albicans*. But it has also long been known that the neutrophil granules contain proteases — protein-degrading enzymes — and two studies revealed that mice deficient in two of these enzymes, elastase and cathepsin G, fared poorly when challenged with a fungal³ or bacterial⁴ infection.

Here, then, was the starting point for the re-examination of the problem by Reeves *et al.*¹. First they established that mice deficient in both enzymes are indeed unable to combat infections by two of the most prevalent pathogens, *C. albicans* and *Staphylococcus aureus* (one of which, they further found, is attacked only by elastase, the other only by cathepsin G). Yet the neutrophils in these animals are otherwise fully functional and display the swift appearance of the reactive oxygen species on ingesting the bacteria. The effect was simulated *in vitro* when protease inhibitors added to normal neutrophils were seen to prevent killing of trapped bacteria.

To find explanations for these effects, Reeves *et al.* undertook a minute examination of the events that unfold in the phagocytic vacuole. In the first place, the release of reactive oxygen species — the 'respiratory burst' — is accompanied by a large disturbance in the internal pH, which rises from 6 to 8. This transcends the release of the predominantly acidic granule contents, for protons are consumed in neutralizing the huge concentration of basic superoxide ions and radicals. In addition, much of the anionic charge is offset by a massive influx of potassium ions through the vacuolar membrane. This, and especially the accumulation of osmotically potent degradation products from the disintegrating microbe, renders the vesicle grossly hypertonic, so that the bacterium within shrinks to half its original volume. If degradation is impeded by the addition of protease inhibitors, swelling is suppressed. After a while, this process supervenes as water slowly enters to counter the osmotic gradient.

What, then, regulates the rate and extent of the water uptake? It seems that the water permeability of the vacuolar membrane is much like that of other membranes, and Reeves *et al.* therefore inferred that expansion of the vacuole might be restricted by a

dense network of cytoskeletal proteins under the membrane, which disperses as the pathogen is digested. Indeed, the recovery of the vacuolar volume could be inhibited by the addition of jasplakinolide, a toxin that stabilizes the actin filaments on which the cohesion of the network depends.

But how do the respiratory burst and the ionic surges in the vacuole induce the process of killing, which superoxide and hydrogen peroxide are by themselves incapable of accomplishing? The key lies in the highly charged matrix within the granules to which, it turns out, the proteases are normally adsorbed. In this state they are inactive and so do not create mayhem in the resting cell. But when the ionic strength inside the vacuole rises, the enzymes are liberated: unleashed, the enzymes are active and attack the microbe, and at the elevated pH their activity is maximal. The membrane-associated network must, of course, prevent the vacuole from swelling by an attendant reversal of the increase in ionic strength, and it is striking that certain microbes (mycobacteria), which evade the innate immune response, can disrupt actin filaments in another cell type⁵. Reeves *et al.* concede that the neutrophil myeloperoxidase does play a significant part in killing. They conjecture that it may protect the proteases themselves from oxidative damage, to which cathepsin G appears to be especially liable. The apparently unwonted complexity of the entire system may have evolved to protect the cell from its own toxins.

The results of this study¹ have implications beyond immunology, for the supposed action of reactive oxygen species on microbes has been taken as a model for their destructive effect on animal tissues. But in particular, it clarifies the action of the innate immune system, the primacy of which was recognized well before the discovery of antibodies and acquired immunity.

That may please the turbulent shade of the despotic Professor Colonel Sir Almroth Wright, founder of the Inoculation Department at St Mary's Hospital Medical School in London. For Wright believed that immunology was the key to the treatment of nearly all important diseases. George Bernard Shaw, in his play *The Doctor's Dilemma*, put into the mouth of Sir Colenso Ridgdon, alias Wright, the insistent injunction to "stimulate the phagocytes".

Walter Gratzer is at the Randall Centre, King's College London, New Hunt's House, Guy's Campus, St Thomas' Street, London SE1 1UL, UK.

e-mail: wbg@helios.raai.kcl.ac.uk

1. Reeves, E. P. *et al.* *Nature* **416**, 291–297 (2002).

2. Babior, B. M., Curnutte, J. T. & Kipnes, R. S. *J. Lab. Clin. Med.* **85**, 235–244 (1975).

3. Tkalecivic, J. *et al.* *Immunity* **12**, 201–210 (2000).

4. Belaouaj, A. *et al.* *Nature Med.* **4**, 615–618 (1998).

5. Guerin, I. & de Chastellier, C. *Infect. Immunol.* **68**, 2655–2662 (2000).

Biosynthesis of an organofluorine molecule

A fluorinase enzyme has been discovered that catalyses carbon–fluorine bond formation.

Although fluorine in the form of fluoride minerals is the most abundant halogen in the Earth's crust, only 12 naturally occurring organofluorine compounds have so far been found¹, and how these are biosynthesized remains a mystery². Here we describe an enzymatic reaction that occurs in the bacterium *Streptomyces cattleya* and which catalyses the conversion of fluoride ion and S-adenosylmethionine (SAM) to 5'-fluoro-5'-deoxyfluoroadenosine (5'-FDA). To our knowledge, this is the first fluorinase enzyme to be identified, a discovery that opens up a new biotechnological opportunity for the preparation of organofluorine compounds.

The rarity of natural fluorinated products contrasts with the identification of about 3,500 naturally occurring halogenated compounds³. The available fluoride is largely insoluble — for example, sea water contains 1.3 p.p.m. fluoride and 19,000 p.p.m. chloride, which may help to explain why fluorine's biochemistry has hardly evolved.

The toxin fluoroacetate is the most ubiqu-

itous of the small class of organofluorine compounds and has been identified in more than 40 plant species from all of the continents apart from Antarctica⁴, but its biosynthetic fluorination pathway has not been clearly defined^{2,5}. Fluoroacetate is also produced by the bacterium *S. cattleya*⁶ when it is grown in culture medium supplemented with fluoride ions.

We investigated this process of enzymatic fluorination by incubating a partially purified protein extract from *S. cattleya* with fluoride ions and SAM, and monitored the reaction by using high-pressure liquid chromatography (HPLC). The initial product of fluorination in *S. cattleya* was shown to be 5'-FDA by reference to a synthetic standard (Fig. 1 legend); the biotransformed compound and standard 5'-FDA co-eluted on HPLC and (after derivatization) had identical molecular masses and ¹⁹F nuclear magnetic resonance (NMR) spectra.

Our crude cell-free protein preparation from *S. cattleya* cells was able to mediate the biotransformation of SAM and fluoride ion all the way to fluoroacetate (Fig. 1), indicating that this organism also contains the necessary enzyme activities to convert 5'-FDA to fluoroacetate. We incubated our synthetic 5'-FDA with the crude protein extract of *S. cattleya* and directly monitored its bioconversion to fluoro-

acetate by ¹⁹F-NMR analysis and HPLC. The fluorination reaction seems to involve a nucleophilic attack by fluoride ion at the C-5' carbon of SAM, generating 5'-FDA and concomitantly displacing L-methionine. The mechanism by which 5'-FDA is metabolized to fluoroacetate remains to be established (Fig. 1), although fluoroacetaldehyde^{7,8} may be the immediate precursor of fluoroacetate.

David O'Hagan*, Christoph Schaffrath*, Steven L. Cobb*, John T. G. Hamilton†, Cormac D. Murphy*

*School of Chemistry and Centre for Biomolecular Sciences, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, UK

e-mail: do1@st-andrews.ac.uk

†Microbial Biochemistry Section, Department of Food Science, The Queen's University of Belfast, Newforge Lane, Belfast BT9 9PX, UK

1. O'Hagan, D. & Harper, D. B. *Nat. Prod. Rep.* **11**, 123–133 (1994).
2. O'Hagan, D. & Harper, D. B. in *Asymmetric Fluoro-organic Chemistry* (ed. Ramchandran, P. V.) 210–225 (Am. Chem. Soc. Symp. Ser. 746, Washington DC, 1999).
3. Gribble, G. W. *Chem. Soc. Rev.* **28**, 335–346 (1999).
4. Hall, R. J. *New Phytol.* **71**, 855–871 (1972).
5. Peters, R. A. & Shorthouse, M. A. *Nature* **216**, 80–81 (1967).
6. Sanada, M. et al. *J. Antibiot.* **39**, 259–265 (1986).
7. Moss, S. J. *Chem. Commun.* 2281–2282 (2000).
8. Murphy, C. D., Moss, J. S. & O'Hagan, D. *Appl. Environ. Microbiol.* **67**, 4919–4921 (2001).
9. Gani, D. & Johnson, A. W. *J. Chem. Soc. Perkin Trans. 1*, 1197–1204 (1982).

Competing financial interests: declared none.

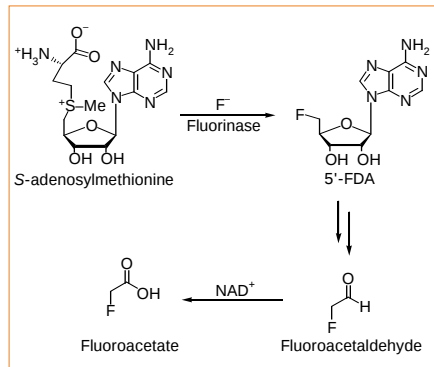


Figure 1 The fluorinase enzyme from *Streptomyces cattleya* mediates the conversion of S-adenosyl-L-methionine (SAM) to 5'-fluoro-5'-deoxyadenosine (5'-FDA). NAD⁺, nicotinic adenine dinucleotide. A cell-free protein extract prepared by sonication of washed *S. cattleya* cells at 4 °C was precipitated with ammonium sulphate (45–60% saturation) and the precipitate dissolved in 50 mM Tris buffer, pH 7.8, before desalting on a HiTrap column (Pharmacia) and gradient elution from a 15 Q anion-exchange column (Pharmacia) with 50 mM Tris, pH 7.8, containing 0–400 mM KCl. Eluted fractions were incubated (960 µl) with SAM (0.4 mM) and KF (10 mM) for 16 h at 26 °C and assayed by high-pressure liquid chromatography (HPLC). 5'-FDA was identified by gas chromatography with mass spectroscopy (GC-MS), HPLC and ¹⁹F-NMR by comparison with a synthetic standard. Synthetic 5'-FDA, prepared by treatment of 2'-O,3'-O-isopropylidene-5'-O-*p*-tosyladenosine with tetrabutylammonium fluoride⁹ and then with dilute sulphuric acid, yielded fluoroacetate when incubated with *S. cattleya* crude protein extract, identified by ¹⁹F-NMR comparison to a reference sample. Details of spectroscopic characterization of standards are available from the authors.

Tumour biology

Herceptin acts as an anti-angiogenic cocktail

Malignant tumours secrete factors that enable them to commandeer their own blood supply (angiogenesis), and blocking the action of these factors can inhibit tumour growth. But because tumours may become resistant to treatments that target individual angiogenic factors by switching over to other angiogenic molecules^{1,2}, a cocktail of multiple anti-angiogenic agents should be more effective. Here we show that herceptin³, a monoclonal antibody against the cell-surface receptor HER2 (for human epidermal growth factor receptor-2; ref. 4), induces normalization and regression of the vasculature in an experimental human breast tumour that overexpresses HER2 in mice, and that it works by modulating the effects of different pro- and anti-angiogenic factors. As a single agent that acts against multiple targets, herceptin, or drugs like it,

may offer a simple alternative to combination anti-angiogenic treatments.

We found that herceptin treatment significantly reduces the diameter and volume, but not the length, of tumour blood vessels compared with those in tumours treated with a control antibody (Fig. 1a, b); vascular permeability is also significantly reduced (8.8 ± 5.7 and $1.6 \pm 1.1 \times 10^{-7}$ cm s⁻¹ at day 15 in the control and herceptin groups, respectively). Blood vessels in the herceptin-treated tumours thus more closely resemble a normal phenotype. Furthermore, tumour growth is slowed (Fig. 1b) and animal survival is significantly extended (Table 1) in response to herceptin treatment. These vascular effects are not dependent on tumour size, as shown by their persistence until the end of the experiment when tumours were largest (6 mm in diameter; Table 1).

To investigate how herceptin achieves this anti-angiogenic effect, we investigated the expression of 23 angiogenesis-related genes by using a gene array. Expression of the pro-angiogenic factors VEGF

(vascular-endothelial growth factor), TGF- α (transforming growth factor- α), Ang-1 (angiopoietin-1) and PAI-1 (plasminogen-activator inhibitor-1) were all reduced, whereas expression of the anti-angiogenic factor TSP-1 (thrombospondin-1) was increased in herceptin-treated tumours relative to control-treated tumours *in vivo* (Table 1). These expression profiles were confirmed by northern-blot analysis.

HER2 signalling is known to control the expression of VEGF (ref. 5) and PAI-1 (ref. 6). HER2 may also affect TGF- α transcripts through interaction with HER1 (ref. 7) and may mediate TSP-1 expression through Ras-like pathways⁵. To our knowledge, no correlation has been reported between HER2 and Ang-1. The vascular effects of herceptin reported here are consistent with an increase in vessel diameter by Ang-1 (ref. 8) and a reduction in vessel diameter by TSP-1 (ref. 9). Vascular permeability is possibly affected by TSP-1 through recruitment of perivascular cells⁹, an effect that would be balanced by the reduction in Ang-1 (ref. 8).

Surprisingly, VEGF expression was reduced by herceptin treatment *in vitro* but not *in vivo* (Table 1), suggesting that host cells may produce compensatory VEGF — an idea that is supported by our immunohistochemical observation of VEGF in host stromal cells as well as in tumour cells (data not shown). Cultured tumour cells also expressed smaller amounts of Ang-1 and PAI-1 and more TSP-1 *in vitro* than tumour tissues did *in vivo*, with the change in PAI-1 expression being greater *in vivo* than *in vitro*. Herceptin treatment may affect PAI-1 from the host by modulation of tumour-derived TGF- α . These results underline the importance of tumour–host interactions in the production of pro- and anti-angiogenic factors.

As the tumour vascular network after herceptin treatment becomes more efficient and more closely resembles normal networks, it might be possible to deliver drugs to previously inaccessible regions¹⁰. However, our findings need to be confirmed in

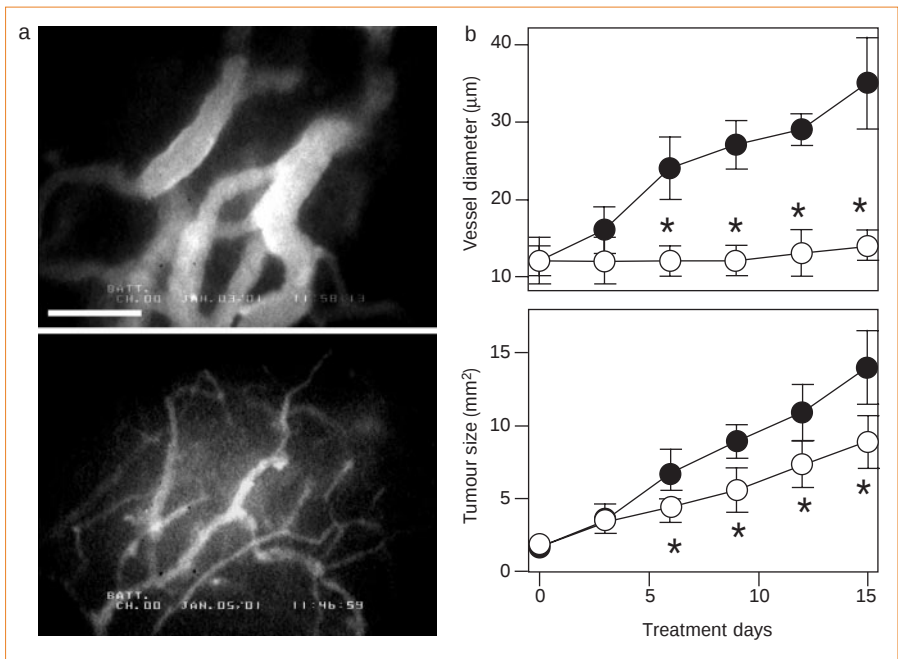


Figure 1 Effects of herceptin on MDA-MB-361HK tumour growth, angiogenesis and gene expression. Tumours were grown in cranial windows in severe combined immunodeficient (SCID) mice and intravital microscopy was carried out as described¹². Established tumours (about 2 mm at treatment day 0) were treated with herceptin or control immunoglobulin G (Genentech), at an intraperitoneal dose of 30 mg kg⁻¹ every 3 days. **a**, Vasculature of control (top) and herceptin-treated (bottom) tumours on treatment day 15. Scale bar, 100 μ m. **b**, Changes in vessel diameter and tumour size during treatment with herceptin (open circles) or control antibody (filled circles) (mean $\pm$ s.d.; $n = 6$; asterisk denotes $P < 0.05$).

other experimental models and also in a clinical setting. We anticipate that, as herceptin primarily affects tumour cells directly and host cells indirectly, angiogenic factors from host cells¹¹ may eventually disrupt this more normal vasculature. Other anti-angiogenic partners (for example, an anti-VEGF agent) may therefore be needed to achieve stable, long-term vessel regression. To identify optimal combinations, the contribution of host stromal cells must also be taken into account, together with other angiogenesis-related factors and the mechanisms that are responsible for controlling tumour–host interactions.

Other inhibitors of signal transduction that have potential as cocktail treatments might block different sets of angiogenic factors. We should eventually be able to obtain angiogenic profiles of individual

tumours and patients, allowing the most appropriate combination of signal-transduction inhibitors to be selected with a view to customizing cancer treatment.

Yotaro Izumi, Lei Xu, Emmanuelle di Tomaso, Dai Fukumura, Rakesh K. Jain

Edwin L. Steele Laboratory, Department of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA

e-mail: jain@steele.mgh.harvard.edu

1. Jain, R. K. & Carmeliet, P. *Sci. Am.* **285**, 38–45 (2001).
2. Fidler, I. J. *J. Natl Cancer Inst.* **93**, 1040–1041 (2001).
3. Slamon, D. J. *et al. N. Engl. J. Med.* **344**, 783–792 (2001).
4. Yarden, Y. & Sliwkowski, M. X. *Nature Rev. Mol. Cell Biol.* **2**, 127–137 (2001).
5. Kerbel, R. S., Vilorio-Petit, A., Klement, G. & Rak, J. *Eur. J. Cancer* **36**, 1248–1257 (2000).
6. Allgayer, H. *et al. J. Clin. Oncol.* **18**, 2201–2209 (2000).
7. Nicolini, G., Miloso, M., Moroni, M. C., Beguinot, L. & Scotto, L. *J. Biol. Chem.* **271**, 30290–30296 (1996).
8. Thurston, G. *et al. Science* **286**, 2511–2514 (1999).
9. Bornstein, P. *J. Clin. Invest.* **107**, 929–934 (2001).
10. Jain, R. K. *Nature Med.* **7**, 987–989 (2001).
11. Fukumura, D. *et al. Cell* **94**, 715–725 (1998).
12. Jain, R. K., Munn, L. L. & Fukumura, D. *Nature Rev. Cancer* (in the press).

Competing financial interests: declared (see online version).

Table 1 Effects of herceptin on an experimental tumour

	Vessel diameter (μ m)	Vessel density (cm cm ⁻²)	Vessel volume (μ m ³ μ m ⁻²)	Permeability (cm s ⁻¹ $\times 10^{-7}$)	Survival (days)
Control	40.4 $\pm$ 6.7	144 $\pm$ 11	25.4 $\pm$ 6.3	8.0 $\pm$ 3.4	20 $\pm$ 3
Herceptin	14.2 $\pm$ 4.1*	181 $\pm$ 70	3.8 $\pm$ 2.3*	2.7 $\pm$ 1.4*	33 $\pm$ 12*

Production of angiogenic factors

	VEGF	TGF- α	Ang-1	PAI-1	TSP-1
Gene array	0.5	0.5	0.6	0.5	4.2
<i>In vivo</i>	0.9	0.5	0.7	0.2	5.3
<i>In vitro</i>	0.3	0.5	0.5	0.7	4.1

The physiological properties of 6-mm tumours (end point) are shown, together with the molecular profiles of angiogenesis-related factors produced by these tumours. MDA-MB-361HK cells were treated *in vitro* with 50 μ g ml⁻¹ herceptin or control human immunoglobulin G for 72 h. Values in the upper section of the table are means $\pm$ s.d. ($n = 6$). In the lower section, messenger RNA abundance for each different factor is shown, normalized to β -actin and expressed as a ratio for herceptin-treated to control-treated tumours. Results were quantified by densitometry.

* $P < 0.05$ (Mann–Whitney U-test). Further details are available from the authors.

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com) or on request from nature@nature.com

The transorientation hypothesis for codon recognition during protein synthesis

Anne B. Simonson & James A. Lake

Molecular Biology Institute, Human Genetics, and MCD Biology, University of California, Los Angeles, California 90095, USA

During decoding, a codon of messenger RNA is matched with its cognate aminoacyl-transfer RNA and the amino acid carried by the tRNA is added to the growing protein chain. Here we propose a molecular mechanism for the decoding phase of translation: the transorientation hypothesis. The model incorporates a newly identified tRNA binding site and utilizes a flip between two tRNA anticodon loop structures, the 5'-stacked and the 3'-stacked conformations. The anticodon loop acts as a three-dimensional hinge permitting rotation of the tRNA about a relatively fixed codon-anticodon pair. This rotation, driven by a conformational change in elongation factor Tu involving GTP hydrolysis, transorients the incoming tRNA into the A site from the D site of initial binding and decoding, where it can be proofread and accommodated. The proposed mechanisms are compatible with the known structures, conformations and functions of the ribosome and its component parts including tRNAs and EF-Tu, in both the GTP and GDP states.

Protein synthesis is a complex, multistep process comprising initiation, elongation, and termination steps¹. Elongation consists of two principal phases. In the first, or decoding phase, a codon of messenger RNA, is matched with its cognate tRNA, and the amino acid carried by the tRNA is added to the growing protein chain. In the second, the messenger RNA and the peptidyl-tRNA are translocated exactly one codon to make room for the next tRNA. In the decoding step, elongation factor Tu (EF-Tu) binds to the ribosome as a ternary complex (EF-Tu·GTP·aminoacyl-tRNA). Upon recognition of the cognate codon, EF-Tu switches conformations and helps to place the aminoacyl-tRNA in the A site. It is thought that the high accuracy of translation results from the combination of an initial decoding step and a second proofreading step following GTP hydrolysis^{2,3}, although alternative models exist⁴.

Although the locations of the A-, P-, and E-tRNA binding sites are known, the site of initial decoding is unclear. The A, P and E sites have been localized on the ribosome using cryoelectron microscopy⁵ and atomic resolution X-ray diffraction^{6,7}. The binding sites of EF-Tu in the presence of the antibiotic kirromycin⁸ and of EF-G in the presence of the antibiotic fusidic acid⁹ have also been mapped, by cryo-electron microscopy. In both mappings the elongation factors are stably bound, but in a modified GDP state. In contrast, the tRNA binding site for the decoding step (the GTP state) has been difficult to map because the ternary complex binds only transiently to programmed ribosomes. Using immuno-electron microscopy, however, the site of ternary complex binding was mapped and a new decoding site, the D site, was identified (see Supplementary Information). This mapping, together with the recent X-ray structures of the ribosome^{6,10–12} and the ternary complex¹³ have allowed us to refine our early working model¹⁴ of decoding during elongation. Here we describe a molecular mechanism for the decoding step: the transorientation hypothesis. In analogy with the EF-G-GTP-dependent translocation of tRNA from the A to the P site, transorientation refers to the EF-Tu-GTP-dependent rotation of tRNA from the D site to the A site.

The transorientation model

In this new model, codon recognition and decoding starts with ternary complex binding to the D site and ends with transfer of the nascent chain to the incoming aminoacyl tRNA positioned in the A site. The model utilizes a conformational switch between two anticodon loop structures, one in which the anticodon is stacked on the 3' side of the anticodon loop (3' stack) and one in which the

anticodon is stacked on the 5' side of the loop (5' stack). Through this switching mechanism, the anticodon remains paired to its cognate codon while the remainder of the tRNA rotates (transorients) from the D site (Fig. 1a) to the A site (Fig. 1b). (This rotation is illustrated in detail in the Supplementary Information).

Transorientation

The 5' stacked conformation of the tRNA anticodon loop is essential to the transorientation hypothesis because it allows the codon-anticodon pair to remain fixed while the aa-tRNA rotates from the D site to the A site. In the 5' stack, anticodon loop nucleotides 32–36 are stacked to continue the 5' strand of the acceptor stem, whereas in the 3' stack nucleotides 34–38 continue the 3' strand of the acceptor stem (Fig. 1e). The 5' stack, first proposed by Fuller and Hodgson¹⁵, has been previously used in models for translocation¹⁶, transorientation¹⁴ and codon evolution¹⁷. No direct X-ray structural evidence for the 5' stacked conformation exists. X-ray diffraction studies of individual tRNAs^{18,19}, of EF-Tu in the ternary complex¹³, and of tRNA synthetase complexes²⁰ have consistently revealed tRNAs in the 3' stacked conformation, although the anticodon loop can be considerably distorted in synthetase structures. A number of important issues, discussed below, must be addressed if transorientation is to be a viable hypothesis.

Can the 5' stacked conformation be built?

Although the 5' stacked tRNA has not been observed in X-ray-derived structures, the 5' stacked structure can be constructed using the bond angles, bond distances, and sugar pucker characteristic of A-form RNAs. A 5' stacked tRNA and a 3' stacked tRNA are shown for comparison docked into the 30S D site in Fig. 1c and d, respectively (oriented as in Fig. 1a). A characteristic of the 5' stacked structure is that the anticodon, shown in magenta, green and blue, stacks against base U₃₃ to continue a helical strand of the anticodon stem.

Where must a 5' stacked tRNA bind in order to read a codon in the A site?

Messenger RNA moves one codon during translocation and shifts little during transorientation, as judged by RNase protection experiments^{21,22}. Thus, the initial decoding site can be determined, in principle, by working backwards using the A site codon orientation as a guide (Fig. 1b). When an aminoacyl tRNA in the 5' stacked conformation is paired to an A-site codon, it fits into a cavity with a complementary shape located on the surface of the 30S ribosomal

subunit. This putative tRNA binding site, the D site, accommodates the anticodon stem loop of the aminoacyl tRNA and neatly positions the variable hypermodified purine, R37, of the 5' stacked structure within a groove formed between parallel RNA helices 34 and 32.

Why cannot a 3' stacked tRNA function in the decoding site?

An aminoacyl tRNA in the 3' stack is docked into the D site in Fig. 1d. Although it fits comfortably into the D site, its anticodon cannot pair with the A site codon. (A tRNA positioned in the EF-Tu-GDP-Phe-tRNA^{Phe}-kirromycin site⁸ mapped by cryo-electron microscopy could pair with the A-site codon, but this antibiotic-induced binding occurs after the decoding step and GTP hydrolysis.) Even when the 3' stacked tRNA is rotated about the long axis of the anticodon stem, it still cannot base pair with the codon, because the anticodon points in the wrong direction, thereby precluding the formation of an antiparallel RNA double helix.

Is there evidence from the literature for a 5' anticodon stack?

Nuclease protection experiments suggest that the 5' stack exists in solution. Nuclease S₁ hydrolyses single-stranded, but not double stranded, nucleic acids and is a sensitive probe of base stacking.

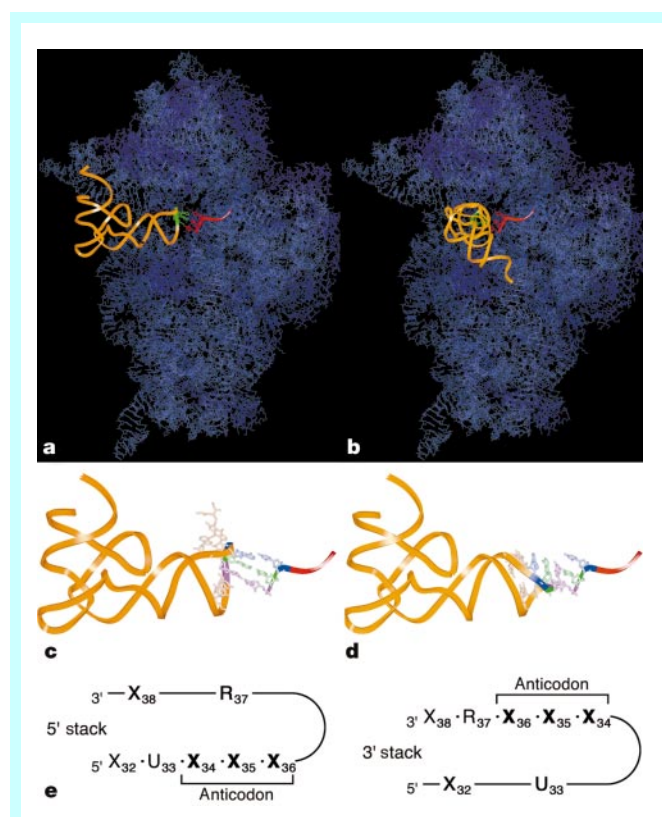


Figure 1 The structures of tRNAs and their orientation on the 30S subunit. **a**, 30S subunit (as viewed for the 50S) with a tRNA in the 5' stacked conformation, docked in the D site. **b**, 30S with a tRNA in the 3' stacked conformation docked in the A site. Pale and dark purple, 30S rRNA and proteins, respectively; orange ribbon, tRNA, with anticodon in green; red ribbon, A site and P site codons. **c**, Close-up view of the 5' stacked tRNA, docked in the D site. **d**, Close-up view of a 3'-stacked tRNA docked into the D site, showing the difficulty in base pairing with the A-site codon. Orange ribbon, 5' stacked tRNA. Anticodon nucleotides 34, 35, and 36, and their complementary codons, are coloured magenta, green, and blue, respectively. **e**, Schematic representation of the 5' and 3' stacked tRNAs. The 30S structure is from ref. 6, and the A- and P-site codons are from ref. 10, except that the A site codon has been modified slightly. Molecular modelling was performed with O, version 7.0.0⁴⁵ and interactive docking was performed with InsightII, V97.5 (Accelrys).

Harada and Dahlberg²³ investigated the cutting of elongator tRNAs at pH 4.5 and found that the primary S₁ cut was at the 3' side of the anticodon, implying a 5' stack for essentially all of the intact molecules. Tal²⁴ independently found similar results at pH 7.5 using a *Neurospora* endonuclease with similar specificity for polynucleotides. Wrede *et al.*²⁵ compared the S₁ cleavage patterns of initiator and elongator tRNAs. For elongator tRNAs they found results similar to those of ref. 23, consistent with the 5' anticodon stack. But for the initiator tRNAs, which enter directly into the P site, and would not need the 5' stacked conformation, they found a unique pattern lacking the characteristic cut expected for the 5' stacked conformation. A recent nuclear magnetic resonance (NMR) comparison of initiator and methionine elongator anticodon loop structures²⁶ showed 'a more flexible structure' in the elongator tRNA than in the initiator tRNA, again consistent with switches between the 3' and 5' stacked conformations in the elongator tRNA. These studies, however, were performed with anticodon stem loops lacking modified bases. Thus there is a significant body of experimental data in the literature consistent with the existence of the 5' stacked anticodon conformation in elongator tRNAs.

What is the mechanism for generating the 5' stack within the ribosome?

In double-stranded B-form DNA, sequences of a pyrimidine (Y) followed by a purine (R), are known as helix breakers since they are characterized by high roll angles and tend to interrupt helices^{27,28}. Similarly, single-stranded A-form RNA helices have decreasing rigidity according to the surface area available for stacking between adjacent bases in the order: 5'RY3' > 5'RR3' > 5'YY3' > 5'YR3', where 5'YR3', the least stable, is a helix breaker. The anticodon is flanked on the 5' side by a universally conserved pyrimidine, U33, and on the 3' side by a purine, R37 (Fig. 1e), which is often hypermodified. As a consequence, flexible 'hot spots', indicated by carets, enclose the anticodon: U33^X34 X35 X36^R37. These 'hot spots' are understood as follows. If the nucleotide at position 34 is a purine, then the dinucleotide Y33R34 is a helix breaker, and if nucleotide 34 is a pyrimidine, then Y33Y34 is the second most flexible dinucleotide. On the other side of the anticodon, if nucleotide 36 of the anticodon is a pyrimidine, then Y36R37 is a helix breaker. However, if position 36 is a purine, then R36R37 is a strong, although not the strongest, dinucleotide. A significant correlation ($P < 0.001$) between the modification of tRNA nucleotide R37 and the presence of R36 suggests that R37 modification could be part of a helix-breaking mechanism that helps convert the 3' stacked tRNA to the 5' stacked conformer (see Supplementary Information).

Other experiments indicate a direct involvement of modified R37 in codon recognition. Carbon and Fleck²⁹ showed that the *su*⁺(A78) suppressor tRNA derived from *E. coli* tRNA^{Gly3}_{GGU/G} requires two changes for suppressor activity (C36 is changed to A36, and A37 is modified to 2-methylthio-N⁶-isopentenyladenosine, ms²i⁶A37). Transfer RNAs with unmodified A37 lacked activity, indicating A37 modification is needed for elongation and is functionally correlated to position 36. Gefter and Russel³⁰ showed that tRNA^{Tyr1}_{UAG} lacking the ms²i⁶A₃₇ modification (that is, anticodon loop A36A37) would neither bind to *E. coli* ribosomes in the presence of its triplet, nor incorporate tyrosine into polypeptides in a cell-free amino-acid incorporating system, whereas the A37-modified tRNA (anticodon loop A36ms²i⁶A37) would. Because the two tRNAs were otherwise unchanged, this directly implicates R37 in tRNA binding to the D site.

Can the 70S accommodate the ternary complex bound in the D site?

The ternary complex docked in the D site fits onto the 70S ribosome extremely well. It is adjacent to proteins L7/L12 which form a conspicuous structural feature of the ribosome: the L7/L12 stalk³¹,

thought to be related to factor binding, but not essential for GTPase activity³². Although this protein complex closely cradles domain I of EF-Tu (Fig. 2), it does so without overlap or other interference. There are no instances of steric hindrance, with one exception. The L11-rRNA complex displays significant steric overlap with EF-Tu and is discussed below.

What is the role of L11 in the transorientation model?

Ribosomal protein L11 and 23S rRNA nucleotides 1,051–1,108 form a well ordered ribonucleoprotein complex that is thought to regulate entry into the ribosome, through switches involving RNA, protein, or both³³. This highly conserved, functionally important complex is a target for antibiotics that disrupt elongation factor function, possibly by preventing EF-G release³⁴.

The structure of the complex in isolation has been solved by X-ray crystallography³⁴. L11 consists of a carboxy-terminal domain, which interacts directly with the 23S rRNA and an amino-terminal domain, which is tethered to the C domain by a flexible linker. Mutations in the L11 gene that result in thiostrepton resistance correspond to protein residues that lie along a narrow cleft between the RNA and the N-terminal domain of the protein, suggesting that thiostrepton binds in the cleft and blocks a putative conformational switch by the L11-rRNA complex^{34,35}.

With the ternary complex docked in the D site, there is extensive clash between EF-Tu domain II and the C-terminal domain of L11, and considerable overlap between the N-terminal domain of L11 and EF-Tu domain III (Fig. 3a). Even with L11 removed, significant overlap still remains between EF-Tu domain III and 23S ribosomal RNA helix 1,057–1,081 (Fig. 3b). Thus the potential clash between EF-Tu and the L11-rRNA complex is predicted to occur independently of whether L11 is present. Given that L11 is not required for

viability in *E. coli*³⁶, the overlap between EF-Tu and helix 1,057–1,081 suggests that the switch may also involve the RNA.

This may represent a mechanism for controlling the access of ternary complex to the ribosome. It is interesting to note that the four RNA helices within 1,051–1,108 are connected without intervening unpaired nucleotides and hence, like a Holliday structure, may switch between either of two conformers, corresponding to alternative helix stacking patterns. A large cavity, sufficiently large to hold L11 and helix 1,057–1,081, is present just above helix 1,087–1,102 and toward the viewer in Fig. 3. If the L11 complex were to act as a gatekeeper, the cavity could accommodate the complex when the gate was open.

What is the role of EF-Tu in transorientation?

EF-Tu is an integral part of the decoding process. It is a three-domain, GTP-binding protein that undergoes dramatic conformational changes upon the hydrolysis of GTP. Domain I binds GTP and is connected to domains II and III through a linker to domain II. Domains II and III form a long cylindrically shaped structure in the ternary complex that contacts the acceptor end and the T-loop of a bound aa-tRNA, respectively. As a result of GTP hydrolysis, domain II moves distally from domain I, and domains II and III rotate 90 degrees about their long axis^{37–39}. Only after GTP hydrolysis does the incoming aminoacyl tRNA enter the A site, suggesting that movements of EF-Tu might physically move aminoacyl tRNAs from the D site to the A site.

The X-ray structures of EF-Tu in the GTP and GDP states suggest that it can transport an aa-tRNA from the D site to the edge of the A site. We note that if EF-Tu domains II and III are allowed to rotate about the linker connecting them to domain I, thereby permitting the tRNA to continuously pair with the anticodon during transor-

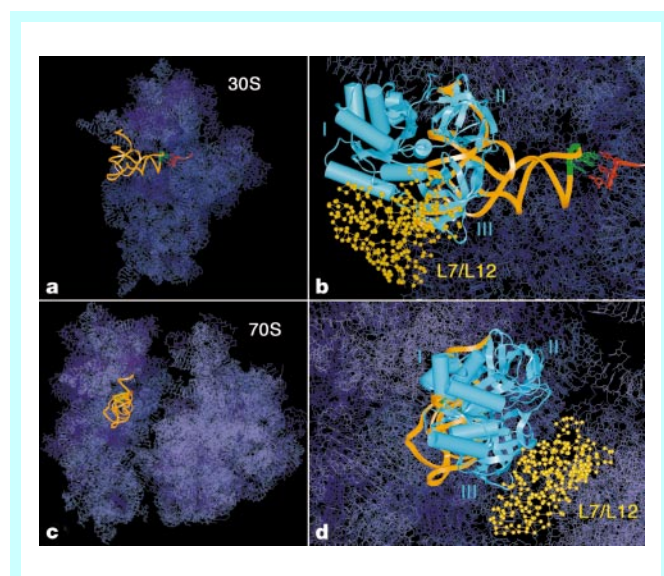


Figure 2 The ternary complex docked into the 70S D site. EF-Tu is coloured cyan, and domains I, II, and III are labelled. The 30S subunit and 70S ribosome are purple, the tRNA is orange, and 50S subunit proteins L7/L12 are shown as yellow ball-and-stick alpha-carbon models. **a, b**, The 30S subunit in the asymmetric projection, as viewed for the 50S subunit; **c, d**, the interface view of the 70S ribosomes. **a** and **c** are designed to orient the reader by illustrating only the aminoacyl tRNA in the D site, whereas **b** and **d** are higher-magnification views of the entire ternary complex. **a**, A 5' stacked tRNA docked in the D site. **b**, A ternary complex docked in the D site. **c**, A 5' stacked tRNA in the D site, showing the 30S and the 50S with L7/L12. **d**, Close-up, ternary complex docked in the D site showing the 30S and the 50S with L7/L12. The binding was tested and analysed using 30S (ref. 6) and 50 (ref. 11) structures aligned with the 70S structure of ref. 10.

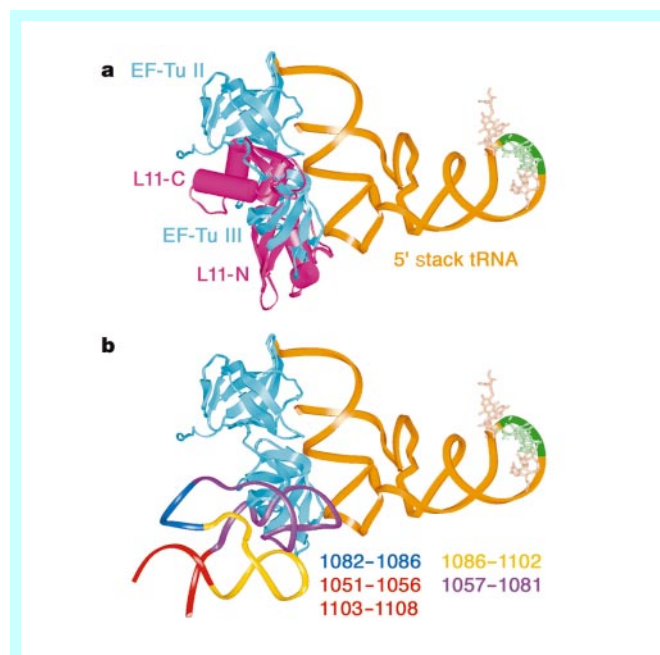


Figure 3 The ternary complex docked in the D site, illustrating steric clash with the 50S L11-rRNA complex. **a**, EF-Tu domains II and III show significant overlap with L11. L11 is shown in magenta. EF-Tu domains are shown in cyan. The tRNA is orange, except the anticodon is green and nucleotides U33 and R37 are pink. L11-C, L11 C-terminal domain; L11-N, L11 N-terminal domain. **b**, EF-Tu domain III and a small portion of EF-Tu domain II exhibit steric overlap in the 1,067 stem-loop of the 23S rRNA. 23S rRNA nucleotide numbering in the *E. coli* system: red, nucleotides 1,051–1,056 (terminal stem); purple, nucleotides 1,057–1,081 (1,067 stem-loop); yellow, nucleotides 1,087–1,102 (1,095 stem-loop); blue, nucleotides 1,082–1,086 (1,082 hairpin).

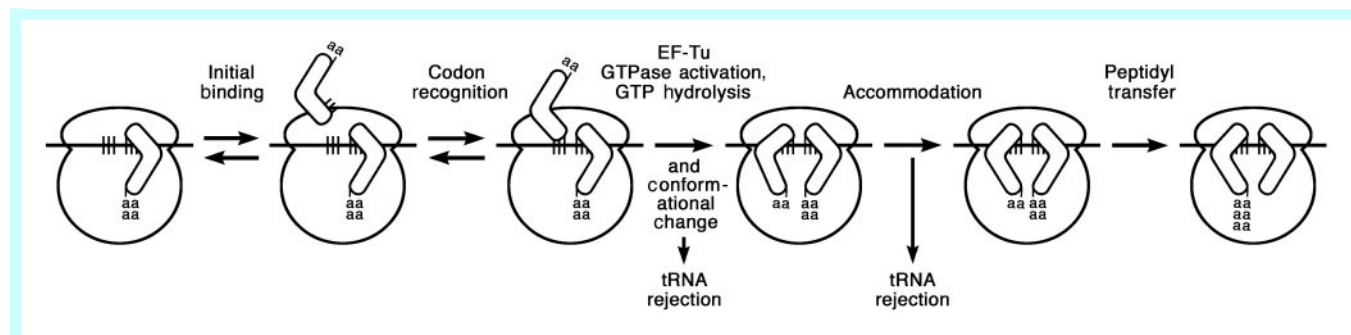


Figure 4 Schematic representation of the ribosome, mRNA and tRNAs during decoding. The figure illustrates the correspondence between the structural transitions

in the transorientation hypothesis and the steps in the kinetic proofreading mechanism observed by ref. 2. See text for details.

ientation, then the motions of EF-Tu are those needed to rotate the tRNA to the A site. The transport occurs without significant clash within EF-Tu or between the EF-Tu-aa-tRNA complex and the 70S ribosome. Furthermore the starting and ending phi and psi angles and the paths connecting them are comfortably within the acceptable range for Ramachandran plots of EF-Tu³⁷. (Details, including illustrations of the starting and ending conformations of the EF-Tu and aa-tRNA, are provided in the Supplementary Information.)

Polekhina *et al.*³⁷ suggested that the function of the EF-Tu Switch I might be to “physically dislocate domains II and III” from domain I. This makes sense if EF-Tu movement powers transorientation, because physical separation of the domains is required to allow the rotation of domains II and III with respect to domain I, without clashing. Similarly, the kirromycin family of antibiotics function by linking domain III to domain I^{40,41}. Again it makes sense that this could inhibit transorientation, since it would prevent the rotation of domains II and III.

How does the transorientation hypothesis fit with proofreading?

Translation requires the recognition of each codon by the correct aminoacyl tRNA. To discriminate accurately⁴² between correct and incorrect aa-RNAs, it is thought that the codon is decoded before, and proofread after, GTP hydrolysis.

The multiple kinetic steps² in codon recognition and proofreading parallel the structural steps in the transorientation hypothesis (Fig. 4). The first step, initial binding, corresponds to the reversible binding of the ternary complex to the decoding site. Codon recognition involves reversible switching of the anticodon conformation to the 5' stack and codon-anticodon pairing. Recognition of the codon triggers a set of irreversible steps involving GTP activation, GTP hydrolysis, and ultimately EF-Tu conformational change. EF-Tu conformational change will test the codon-anticodon interaction, so it may provide an opportunity to proofread the codon (tRNA rejection). Similarly, the process of accommodation during which the aminoacyl tRNA leaves EF-Tu-GDP and enters the A site may also provide an opportunity for proofreading as suggested by studies of antibiotic involvement in misreading and fidelity⁷. Finally, peptidyl transfer irreversibly completes the transorientation minicycle.

Discussion

Transorientation is a very specific hypothesis, and hence there are many ways in which it may be falsified. This makes it eminently testable, that is, it has high popperian verifiability. Our initial tests indicate that the transorientation hypothesis is consistent with many details of ribosome structure. A tRNA in the 5' stacked conformation fits nearly perfectly into its complementary D site. The observed S1 nuclease digests of tRNAs very closely match those predicted by a 5' stacked loop conformation. The parallels observed

between the structural details of transorientation and the energetics of codon recognition and proofreading are also surprising. It is remarkable that the conformational changes between EF-Tu-GTP and EF-Tu-GDP are such that they can direct an incoming aminoacyl tRNA from the D site to the A site.

The models presented here should not be taken as being comparable to X-ray or NMR structures; they are not. Rather, they should be regarded as the first steps taken toward modelling the functioning of a fascinating machinery, which will surely occupy many of us for several decades.

Finally, it is interesting that transorientation is dominated and shaped by the structure and the movements of the tRNA molecules. The ribosome acts as a facilitator, but the tRNA is essential to the mechanism. We can imagine how protein synthesis could have evolved, perhaps beginning from only mRNAs and tRNAs. As such, this mechanism is consistent with the view that protein synthesis may have originated in a pre-existing RNA world^{43,44}. □

Received 25 July 2001; accepted 25 January 2002.

- Garrett, R. A. *et al.* *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions* (ASM, Washington DC, 2000).
- Pape, T., Wintermeyer, W. & Rodnina, M. V. Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the *E. coli* ribosome. *EMBO J.* **17**, 7490–7497 (1998).
- Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl Acad. Sci. USA* **71**, 4135–4139 (1974).
- Nierhaus, K. H. The allosteric three-site model for the ribosomal elongation cycle: features and future. *Biochemistry* **29**, 4997–5008 (1990).
- Agrawal, R. K. *et al.* Direct visualization of A-, P-, and E-site transfer RNAs in the *E. coli* ribosome. *Science* **271**, 1000–1002 (1996).
- Wimberly, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339 (2000).
- Carter, A. P. *et al.* Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407**, 340–348 (2000).
- Stark, H. *et al.* Visualization of elongation factor Tu on the *Escherichia coli* ribosome. *Nature* **389**, 403–406 (1997).
- Agrawal, R. K., Penczek, P., Grassucci, R. A. & Frank, J. Visualization of elongation factor G on the *Escherichia coli* ribosome: the mechanism of translocation. *Proc. Natl Acad. Sci. USA* **95**, 6134–6138 (1998).
- Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).
- Ban, N. *et al.* The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
- Schluzien, F. *et al.* Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**, 615–623 (2000).
- Nissen, P. *et al.* Crystal structure of the ternary complex of the Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* **270**, 1464–1472 (1995).
- Lake, J. A. Aminoacyl-tRNA binding at the recognition site is the first step of the elongation cycle of protein synthesis. *Proc. Natl Acad. Sci. USA* **74**, 1903–1907 (1977).
- Fuller, W. & Hodgson, A. Conformation of the anticodon loop in tRNA. *Nature* **215**, 817–821 (1967).
- Woese, C. Molecular mechanics of translation: a reciprocating ratchet mechanism. *Nature* **226**, 817–820 (1970).
- Crick, F. H. C., Brenner, S., Klug, A. & Piezenik, G. Speculation on the origin of protein synthesis. *Orig. Life* **7**, 389–397 (1976).
- Ladner, J. E. *et al.* Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution. *Proc. Natl Acad. Sci. USA* **72**, 4414–4418 (1975).
- Quigley, G. J., Seeman, N. C., Wang, A., Suddath, F. L. & Rich, A. Yeast phenylalanine transfer RNA: Atomic coordinates and torsion angles. *Nucleic Acids Res.* **2**, 2329–2349 (1975).

20. Eiler, S. *et al.* Synthesis of aspartyl-tRNA^{Asp} in *Escherichia coli*—a snapshot of the second step. *EMBO J.* **18**, 6532–6541 (1999).
21. Gupta, S. L., Waterson, J., Sopor, M. L., Weissman, S. M. & Lengyel, P. Movement of the ribosome along the messenger ribonucleic acid during protein synthesis. *Biochemistry* **10**, 4410–4421 (1971).
22. Thach, S. S. & Thach, R. E. Translocation of messenger RNA and accommodation of fMet-tRNA. *Proc. Natl Acad. Sci. USA* **68**, 1791–1795 (1971).
23. Harada, F. & Dahlberg, J. E. Specific cleavage of tRNA by nuclease S1. *Nucleic Acids Res.* **2**, 865–871 (1975).
24. Tal, J. The cleavage of transfer RNA by a single strand specific endonuclease from *Neurospora crassa*. *Nucleic Acids Res.* **2**, 1073–1082 (1975).
25. Wrede, P., Woo, N. H. & Rich, A. Initiator tRNAs have a unique anticodon loop conformation. *Proc. Natl Acad. Sci. USA* **76**, 3289–3293 (1979).
26. Schweisguth, D. C. & Moore, P. B. On the conformation of the anticodon loops of initiator and elongator methionine tRNAs. *J. Mol. Biol.* **267**, 505–519 (1997).
27. Klug, A. *et al.* A hypothesis on a specific sequence-dependent conformation of DNA and its relation to the binding of the lac-repressor protein. *J. Mol. Biol.* **131**, 669–680 (1979).
28. Dickerson, R. E. & Chiu, T. K. Helix bending as a factor in protein/DNA recognition. *Biopolymers* **44**, 361–403 (1997).
29. Carbon, J. & Fleck, E. W. Genetic alteration of structure and function in glycine transfer RNA of *Escherichia coli*: mechanism of suppression of the tryptophan synthetase A78 mutation. *J. Mol. Biol.* **85**, 371–391 (1974).
30. Gefter, M. L. & Russell, R. L. Role modifications in tyrosine transfer RNA: a modified base affecting ribosome binding. *J. Mol. Biol.* **39**, 145–157 (1969).
31. Strycharz, W. A., Nomura, M. & Lake, J. A. Ribosomal proteins L7/L12 localized at a single region of the large subunit by immune electron microscopy. *J. Mol. Biol.* **126**, 123–140 (1978).
32. Briones, E., Briones, C., Remacha, M. & Ballesta, J. P. The GTPase center protein L12 is required for correct ribosomal stalk assembly but not for *Saccharomyces cerevisiae* viability. *J. Biol. Chem.* **273**, 31956–31961 (1998).
33. Cundliffe, E. in *Structure, Function and Genetics of Ribosomes* (eds Hardesty, B. & Kramer, G.) 586–604 (Springer, New York, 1986).
34. Wimberly, B. T. *et al.* A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell* **97**, 491–502 (1999).
35. Markus, M. A. *et al.* High resolution solution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA. *Nature Struct. Biol.* **4**, 70–77 (1997).
36. Stoeffler, G., Cundliffe, E., Stoeffler-Meilicke, M. & Dabbs, E. R. Mutants of *Escherichia coli* lacking ribosomal protein L11. *J. Biol. Chem.* **255**, 10517–10522 (1980).
37. Polekhina, G. *et al.* Helix unwinding in the effector region of elongation factor EF-Tu GDP. *Structure* **4**, 1141–1151 (1996).
38. Abel, K. *et al.* An α to β conformational switch in EF-Tu. *Structure* **4**, 1153–1159 (1996).
39. Song, H. *et al.* Crystal structure of intact elongation factor EF-Tu from *Escherichia coli* in GDP conformation at 2.05 Å resolution. *J. Mol. Biol.* **285**, 1245–1256 (1999).
40. Abdulkarim, F., Liljas, L. & Hughes, D. Mutations to kirromycin resistance occur in the interface of domains I and III of EF-Tu-GTP. *FEBS Lett.* **352**, 118–122 (1994).
41. Vogeley, L., Palm, G. J., Mesters, J. R. & Hilgenfeld, R. Conformational change of elongation factor Tu induced by antibiotic binding: crystal structure of the complex between EF-Tu:GDP and aurodox. *J. Biol. Chem.* **276**, 17149–17155 (2001).
42. Kurland, C. G. Translational accuracy and the fitness of bacteria. *Annu. Rev. Genet.* **26**, 29–50 (1992).
43. Noller, H. E., Hoffarth, V. & Zimniak, L. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**, 1416–1419 (1992).
44. Zaug, A. J. & Cech, T. R. The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science* **231**, 470–475 (1986).
45. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank D. Eisenberg and R. Dickerson for advice and for access to graphics terminals and software, A. Maris and D. Cascio for help, M. Rivera, R. Jain and J. Moore for advice, and M. Kowalczyk for the art work. This work was supported by grants to J.A.L. from the National Institutes of Health, the Astrobiology Institute, the National Science Foundation, and the Department of Energy.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.L. (e-mail: Lake@mbi.ucla.edu) or A.B.S. (e-mail: asimonso@ucla.edu). The following crystal structures were used in this work, with Protein Data Bank accession numbers shown in parentheses: 50S ribosomal subunit (1FFK), 30S ribosomal subunit (1FJF), 70S ribosome (1GIX, 1GIY), ternary complex (1TTT), phenylalanyl-tRNA (1EVV), Ef-Tu-GDP (1EFC), L11-RNA complex (1MMS). Coordinates for the 5'-stacked tRNA (1KS1), and for the ternary complex docked in the D site (1L1U) have been deposited in the Protein Data Bank.

Diurnal modulation of pacemaker potentials and calcium current in the mammalian circadian clock

Cyriel M. A. Pennartz, Marcel T. G. de Jeu, Nico P. A. Bos, Jeroen Schaap & Alwin M. S. Geurtsen

Graduate School Neurosciences Amsterdam, Netherlands Institute for Brain Research, Meibergdreef 33, 1105 AZ, Amsterdam, The Netherlands

The central biological clock of the mammalian brain is located in the suprachiasmatic nucleus. This hypothalamic region contains neurons that generate a circadian rhythm on a single-cell basis. Clock cells transmit their circadian timing signals to other brain areas by diurnal modulation of their spontaneous firing rate. The intracellular mechanism underlying rhythm generation is thought to consist of one or more self-regulating molecular loops, but it is unknown how these loops interact with the plasma membrane to modulate the ionic conductances that regulate firing behaviour. Here we demonstrate a diurnal modulation of Ca^{2+} current in suprachiasmatic neurons. This current strongly contributes to the generation of spontaneous oscillations in membrane potential, which occur selectively during daytime and are tightly coupled to spike generation. Thus, day–night modulation of Ca^{2+} current is a central step in transducing the intracellular cycling of molecular clocks to the rhythm in spontaneous firing rate.

It has long been known that neurons in the suprachiasmatic nucleus (SCN) transmit circadian output to other brain areas by diurnal modulation of their spontaneous discharge frequency^{1–5}. When firing activity of SCN neurons is monitored in cultures of dissociated cells, many of them express a circadian rhythm that is asynchronous with that of other cells, indicating the cell-autonomous nature of the oscillator⁴. The molecular loops operating within clock cells involve transcription, translation and negative protein-mediated feedback on gene expression^{5,6}. Much less is known about the signals by which the core loop communicates with ionic channels and transporters in the plasma membrane, which directly regulate membrane excitability. Ionic channels may be under direct transcriptional control of the core loop or, alternatively, may be regulated by as-yet unidentified clock-controlled genes^{5,6}. A principal step in the elucidation of this problem is to dissect the ionic mechanisms by which the circadian message from the molecular clock is transduced into bioelectric output. This, however, is not a simple problem. First, there are many ionic currents in the SCN that may be targeted by the molecular clock^{7–10}, yet their contribution to spontaneous firing is largely unknown. Second, ionic currents in SCN neurons can be monitored by patch-clamp recordings for several hours¹⁰, which is, however, too short to study circadian modulation within a single neuron. This technical limitation necessitates groupwise comparisons between neurons recorded during different circadian phases. This consideration, combined with the finding that clock cells become desynchronized in dissociated-cell cultures⁴, led us to use acutely prepared brain slices in studying diurnal modulation of ionic conductances. Before investigating specific conductances, however, we assessed day–night differences in integrated electrophysiological behaviour of SCN cells, which can be optimally studied with the use of perforated patch recordings in current clamp mode¹⁰.

Oscillations in membrane potential

We first verified whether our preparation was capable of expressing a circadian rhythm in firing. Indeed, the spontaneous firing rate (SFR) was higher in neurons recorded during day (8.3 ± 0.6 Hz, mean $\pm$ s.e.m.; $n = 51$) than in the night (2.5 ± 0.5 Hz, $n = 36$) ($P < 1 \times 10^{-6}$; compare with refs 2, 4). We next asked whether diurnal differences in membrane properties would be detectable in the absence of firing activity. After spiking was abolished by

tetrodotoxin (TTX; $1 \mu\text{M}$), we observed two dramatic day–night differences: day cells ($n = 41$) were strongly and tonically depolarized with respect to night cells ($n = 21$) (compare with ref. 10) and exhibited membrane potential oscillations at a frequency of 2–7 Hz that were absent in night cells (Fig. 1). The time-averaged membrane potentials were -43 ± 1 and -55 ± 1 mV for the day and night phases, respectively ($P < 1 \times 10^{-6}$). This difference was associated with an enhanced input resistance during the day ($2.3 \pm 0.1 \text{ G}\Omega$) compared with the night ($1.3 \pm 0.2 \text{ G}\Omega$) ($P < 5 \times 10^{-5}$). The mean frequency and peak-to-valley amplitude of the oscillations were respectively 4.0 ± 0.3 Hz and 14 ± 1 mV during the day ($n = 41$), whereas clear oscillations were only rarely detectable during the night (mean frequency, 0.4 ± 0.2 Hz; amplitude, 1 ± 1 mV; $n = 21$) (both $P < 1 \times 10^{-6}$). The fraction of day cells showing oscillations (37 of 41 cells; 90%) was significantly larger than that during the night (3 of 21 cells; 15%) ($P < 5 \times 10^{-9}$). We quantified the relative strength of the oscillations by spectral analysis of membrane potential traces and by computing the ratio of the peak in the power spectrum in the 2–7-Hz frequency band over the power averaged across the 0.5–1.0-Hz band ('peak/basal ratio'). This ratio was dramatically larger during the day ($41.4 \pm 7.8 \text{ mV}^2$) than the night ($2.5 \pm 1.5 \text{ mV}^2$) ($P < 1 \times 10^{-6}$). The oscillations displayed a distinct voltage dependence: they became smaller and less regular with depolarization above or hyperpolarization below the range of resting potentials (Fig. 1c). The paucity of oscillations in night cells could not be ascribed to their more hyperpolarized state, because depolarization by current injection failed to induce sustained oscillatory behaviour ($n = 21$; Fig. 1b). The TTX-induced depolarization in day cells (Fig. 1d) could not be attributed to a gradual deterioration of cell quality, because it was selectively found in day but not night cells and because the membrane potential recovered after prolonged TTX washout in all 12 day cells tested (Fig. 1a).

We next examined the ionic mechanism underlying the oscillations occurring during daytime. When we replaced Ca^{2+} with Mg^{2+} in a bathing medium containing TTX, the oscillation was strongly and reversibly reduced, as quantified by the peak/basal ratio ($P < 0.05$, $n = 6$; Fig. 2a). Furthermore, the L-type Ca^{2+} channel blocker nimodipine^{11–13} ($2 \mu\text{M}$) virtually abolished the oscillations ($P < 0.02$, $n = 7$; Fig. 2b). Neither treatment affected the time-averaged membrane potential or input resistance

(Fig. 2e, f). Indeed, in the presence of both TTX and nimodipine the membrane potential during the day (-45 ± 2 mV, $n = 7$) remained significantly different from the night (-54 ± 2 mV, $n = 5$) ($P < 0.05$), and the same applied to the input resistance (day, 2.0 ± 0.2 G Ω ; night, 1.2 ± 0.2 G Ω ; $P < 0.05$). The suppression of oscillations was also clearly present, albeit less strong, when using a tenfold lower dose of nimodipine ($0.2 \mu\text{M}$; $n = 3$). In contrast, the oscillations were not affected by the following: $50 \mu\text{M}$ Ni^{2+} , a preferential low-voltage-activated Ca^{2+} channel blocker^{7,11,12,14,15} ($n = 3$); the N-type Ca^{2+} channel blocker ω -conotoxin GVIA^{7,11,13} ($1 \mu\text{M}$; $n = 4$); the P/Q-type Ca^{2+} channel blocker ω -agatoxin IVB¹⁶

($0.2 \mu\text{M}$; $n = 4$); and an enhancement of the TTX dose to $10 \mu\text{M}$ ($n = 3$). First, these results indicate a major role for L-type Ca^{2+} channels in mediating the oscillations. These channels are defined by their relatively long-lasting currents evoked by depolarizing voltage steps^{11–14}. Second, the oscillations and tonically depolarized membrane potential during the day are dissociable and should thus be attributed to different ionic mechanisms.

L-type calcium current

Led by the observation that no sustained oscillations could be induced in night cells, we predicted that Ca^{2+} currents display a larger amplitude during the day than the night. To test this, we recorded current–voltage relationships in voltage-clamp mode. Thresholds of Ca^{2+} currents were situated at -60 to -50 mV in most cells, with low-threshold (T-type) current being absent or of minor amplitude (Fig. 3; holding potential -90 mV). The decay of Ca^{2+} current was marked by a fast and slow component. Their time constants were not significantly different for the day (τ_1 , 15 ± 1 ms; τ_2 , 177 ± 13 ms; $n = 20$) and night (τ_1 , 16 ± 1 ms; τ_2 , 183 ± 27 ms; $n = 21$) (quantified at steps to -20 mV). However, the peak amplitudes were significantly larger during the day (-400 ± 42 pA at -20 mV) than night (-241 ± 30 pA) ($P < 0.01$; Fig. 3d). In addition, the amplitude of the slowly decaying component, as measured 150 ms after step onset, was significantly larger in day cells (-145 ± 19 pA) than in night cells

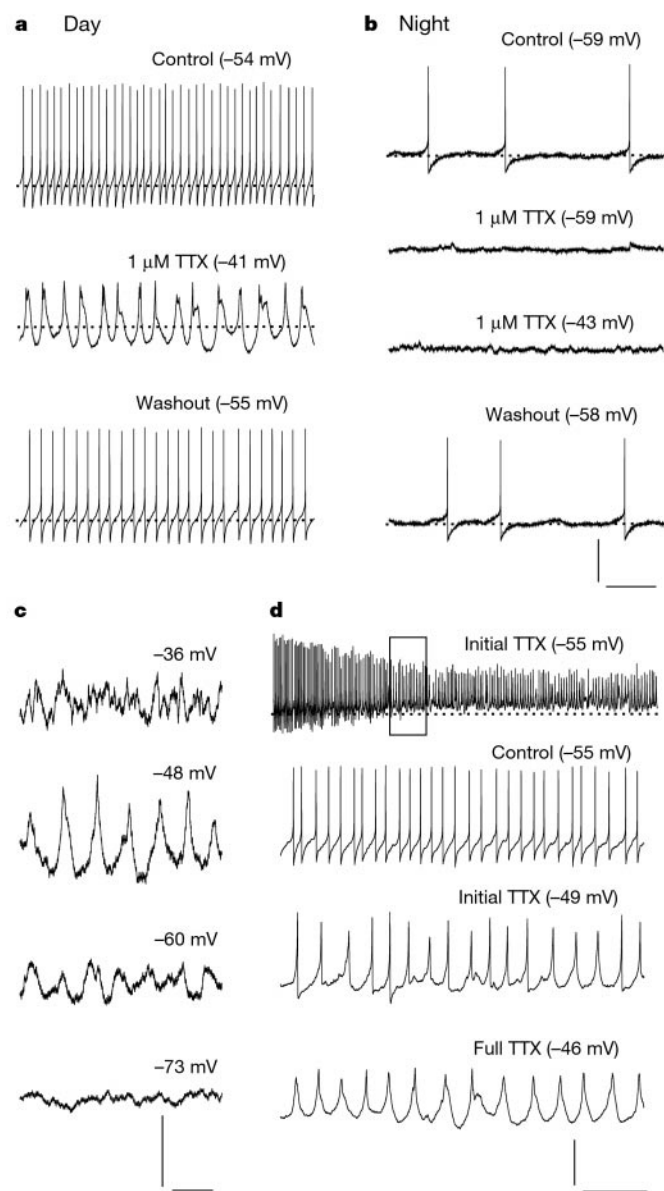


Figure 1 Properties of membrane potential oscillations in SCN. **a**, Day: TTX revealed oscillations and tonically depolarized the membrane. Dotted lines, the time-averaged membrane potential values. **b**, Night: TTX failed to uncover oscillations or depolarize the cell. In the third trace from top, the cell was depolarized by constant current injection. **c**, Voltage dependence of oscillations during TTX. Membrane potential was varied by constant current injection. **d**, Spikes were closely associated with the peak phase of oscillations. The top trace presents the phase of TTX wash-in. Dotted line, membrane potential before TTX. The inset is enlarged in the third trace from the top and represents the mid-phase of TTX wash-in. Calibration bars: **a**, **b**, 40 mV, 500 ms, except middle trace in **a**, 20 mV; **c**, 10 mV, 500 ms; **d**, top, 30 mV, 5000 ms; **d**, second trace, 40 mV, 500 ms; **d**, bottom traces, 20 mV, 500 ms.

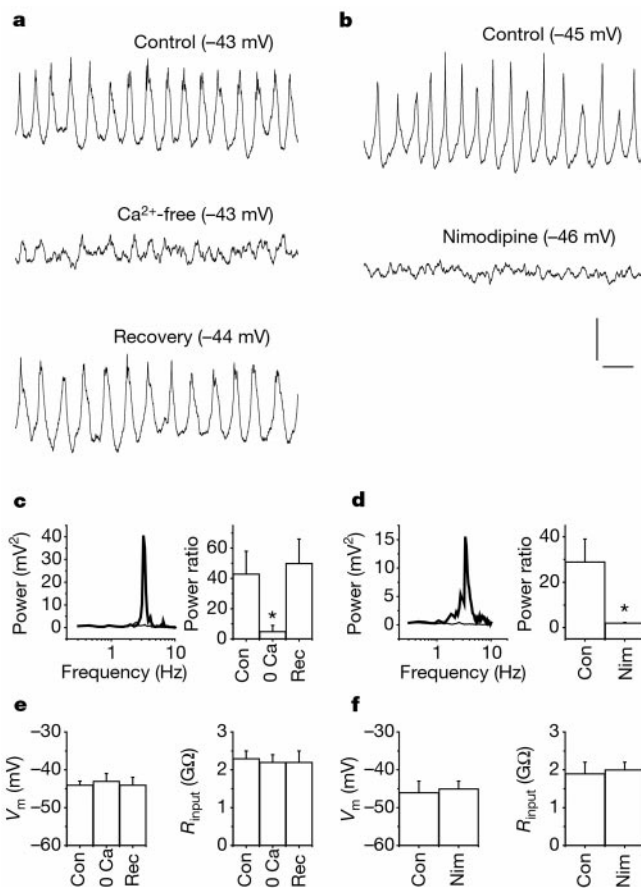


Figure 2 Role of Ca^{2+} channels in mediating daytime oscillations. **a**, Application of Ca^{2+} -free medium reversibly suppressed the oscillations. **b**, Nimodipine ($2 \mu\text{M}$) also abolished the oscillations. Calibration bars (**a** and **b**), 10 mV, 500 ms. **c**, Power spectrum computed over 30-s traces (left). Thick and thin curves, control and Ca^{2+} -free conditions. Peak/basal ratios of spectral components in control (Con), Ca^{2+} -free medium (0 Ca; asterisk, $P < 0.05$) and recovery (Rec) (right). **d**, As for **c**, but for nimodipine (Nim) (asterisk, $P < 0.02$). **e**, **f**, Lack of significant changes in membrane potential (V_m) and input resistance (R_{input}) in Ca^{2+} -free medium (**e**) or nimodipine (**f**).

(-72 ± 11 pA) ($P < 0.002$). A similar day–night difference was found when using a holding potential of -70 mV, making a contribution by T-type Ca^{2+} channels unlikely^{7,11,12,14}. These differences could not be attributed to a difference in cell capacitance (day, 10.3 ± 1.0 pF; night, 8.9 ± 0.8 pF; not significant) or in the voltage dependence of current inactivation. In line with the suppressive effect of nimodipine on daytime oscillations, we found a larger effect of this compound ($2 \mu\text{M}$) on Ca^{2+} currents during the day than the night. Nimodipine strongly reduced the peak of Ca^{2+} current during the day ($P < 0.01$, $n = 9$), whereas the effect at night was modest ($P < 0.05$, $n = 9$) (Fig. 3e–k). When the amount by which nimodipine reduced the peak current was calculated as a percentage of control, the reduction was stronger for the day ($38 \pm 5\%$) than the night ($19 \pm 3\%$) ($P < 0.02$). Similarly, the slowly decaying component was attenuated more strongly during the day ($35 \pm 6\%$) than the night ($14 \pm 4\%$) ($P < 0.02$). Notably, the effect of nimodipine was already visible at voltage levels as

low as -50 mV (Fig. 3f; $n = 5$). During nimodipine, a significant day–night difference in Ca^{2+} current remained visible (day, -236 ± 44 pA; night, -122 ± 24 pA; $P < 0.01$; peak values at -20 mV).

Regulation of spontaneous firing

We wondered about the relevance of the L-type Ca^{2+} current and the oscillations to which it contributes for the regulation of spontaneous firing in SCN neurons. A first indication for their importance was derived from the initial wash-in phase of our TTX experiments (Fig. 1d; $n = 9$), when spikes were not yet fully blocked but the underlying oscillations were already apparent. Attenuated spikes were consistently observed during the peak but not valley phase of the oscillations. Second, the spontaneous firing rate of day neurons in control medium was positively correlated to their oscillation frequency as measured during TTX (correlation coefficient $r = 0.69$, $P < 5 \times 10^{-6}$, $n = 41$; for the overall population of cells, $r = 0.82$, $P < 1 \times 10^{-6}$, $n = 62$). Third, when we tested the effect of nimodipine on spontaneous firing, we found that the firing rate, spike amplitude and amplitude of the spike afterhyperpolarization (compare with ref. 17) decreased in day ($n = 6$; all $P < 0.05$) but not night cells ($n = 6$) (Fig. 4). These findings suggest that L-type Ca^{2+} current contributes to robust patterning of spikes and ensuing afterhyperpolarizations. Accordingly, we propose that the oscillatory potentials emerging during TTX treatment are considered pacemaker potentials in the sense that they drive and temporally organize spontaneous firing in the presence of functioning Na^+ channels. The effect of nimodipine on spike amplitude may be attributed to a relative lack of Na^+ channel deinactivation due to diminished afterhyperpolarizations between successive spikes, in combination with the tonically depolarized membrane potential during daytime (compare with ref. 7).

Discussion

Our results demonstrate a robust day–night difference in L-type Ca^{2+} current in rat SCN, and thus constitute, to our knowledge, the first evidence for diurnal regulation of intrinsic ionic currents in mammals. L-type Ca^{2+} current contributes to robust, high-frequency firing during daytime and is therefore likely to function as a crucial intermediate between the intracellular molecular clock and the bioelectric output of the SCN, which drives and synchronizes circadian rhythms in the brain and body. Previously, a pioneering study in retinal cells of the mollusc *Bulla gouldiana*¹⁸ reported circadian modulation of a K^+ current regulating the basal membrane potential. With regard to vertebrates, a day–night modulation was shown to occur in a mixed cationic current in cultured chick pineal cells¹⁹, but a causal link between this current and spike activity or melatonin release was not demonstrated, and the avian pineal system differs from the mammalian circadian system in a fundamental way²⁰. Our findings suggest how diurnal modulation of Ca^{2+} current contributes to the circadian rhythm in firing rate: Ca^{2+} current mediates the rising phase of 2–7-Hz oscillations and thereby provides an excitatory drive for firing, whereas the falling phase is mediated by a K^+ current and may support Na^+ channel deinactivation. Additional data (not shown here) suggest that this K^+ current is sensitive to tetraethylammonium but not to conventional blockers of Ca^{2+} -dependent K^+ current. Our findings do not imply that this current is also diurnally modulated, nor do they exclude involvement of other currents in oscillatory behaviour or in spontaneous firing⁸.

The high amplitude and frequency of SCN daytime oscillations makes them unique relative to oscillatory phenomena in other types of neurons, although dopamine cells may show similar behaviour under particular conditions²¹. The dihydropyridine sensitivity of the oscillations is consistent with the sustained nature of L-type current and its relatively negative voltage range of activation (Fig. 3f), also reported in several other preparations^{12,22,23} (but see ref.

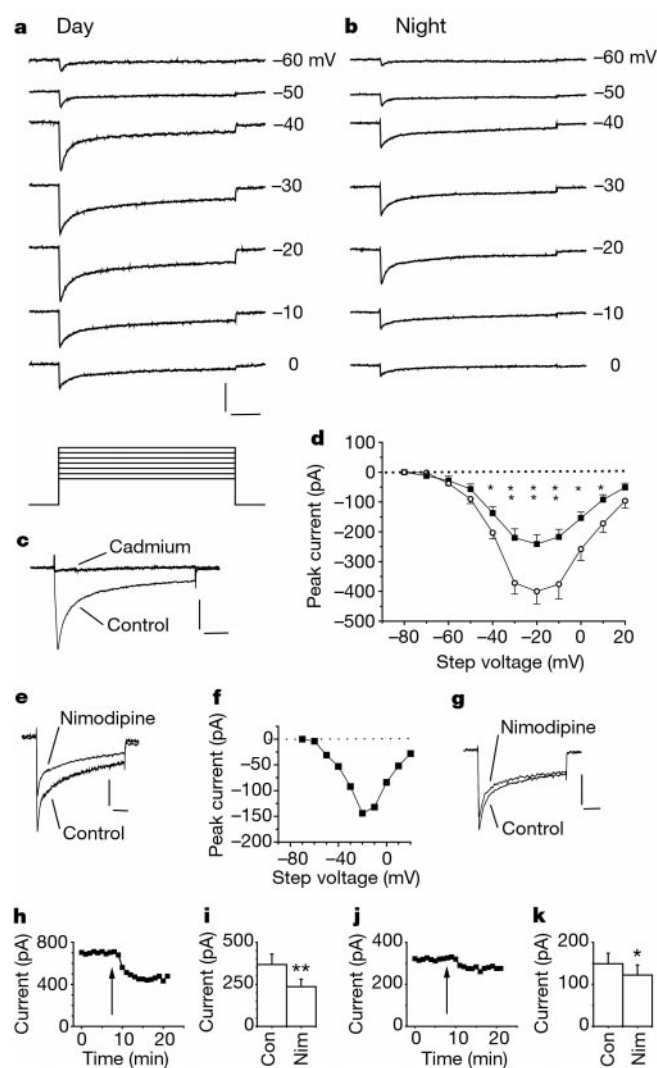


Figure 3 Diurnal modulation of Ca^{2+} current in SCN. **a, b**, Activation curves for day and night cells. **c**, Block of Ca^{2+} current by $200 \mu\text{M}$ Cd^{2+} . **d**, Averaged activation curves for day (open circles) and night (filled squares). Asterisk, $P < 0.05$; double asterisk, $P < 0.01$. **e, g**, Effect of nimodipine ($2 \mu\text{M}$) on Ca^{2+} current during the day (**e**) and night (**g**). **f**, Activation curve of the nimodipine-sensitive component of Ca^{2+} current. **h**, Time course of peak current in the same experiment as **e**. Arrow, onset of nimodipine application. **i**, Effect of nimodipine on peak Ca^{2+} current during day. **j, k**, As for **h** and **i**, but during the night. Calibration bars: **a, b**, 200 pA, 50 ms; **c, e**, 150 pA, 50 ms; **g**, 100 pA, 50 ms.

14). Periodic Ca^{2+} entry through L-type channels is likely to have a significant impact on intracellular Ca^{2+} homeostasis. Indeed it has been reported that the intracellular Ca^{2+} concentration in SCN neurons is higher during the day than night²⁴. Periodic Ca^{2+} entry may alter the activation state of Ca^{2+} -dependent enzymes involved in phase shifting²⁵ and is likely to affect gene expression^{26,27}.

In parallel with the Ca^{2+} current modulation, our evidence substantiates the existence of a second, functionally dissociable ionic mechanism underlying the membrane depolarization and enhanced input resistance during daytime (see also ref. 10). The importance of this mechanism is underscored by the day–night difference in excitability that remains clearly visible in the presence of nimodipine (Fig. 4a, b). It probably involves a daytime reduction of a tonically active K^+ current that is central in maintaining the membrane potential *per se* and is sensitive to internal Cs^+ but insensitive to tetraethylammonium (data not shown). Regardless of the precise molecular identity of this current, it can be inferred that at least two ionic mechanisms are subject to diurnal modulation. An as-yet unidentified nimodipine-insensitive component of Ca^{2+} current is likely to be modulated as well (Fig. 3i, k). Thus, the intracellular molecular clock exerts divergent effects on multiple ion channel and/or transporter species, acting in concert to generate the circadian rhythm in firing rate. Now that a well-defined ion channel in SCN has been identified as a target for diurnal modulation, the

underlying mechanism can be investigated. The day–night difference in amplitude of the L-type Ca^{2+} current may result from a diurnal change in transcription or translation⁵ of Ca^{2+} channel subunits, or in post-translational modifications such as phosphorylation^{28,29} or binding of cofactors³⁰. An important constraint on such changes is defined by our finding that neither the decay kinetics nor the voltage dependence of Ca^{2+} current were markedly altered. □

Methods

Slice preparation

Male Wistar rats (150–300 g) were subjected to a 12:12 h light:dark cycle for at least 3 weeks before use. Rats used for subjective night recordings were housed in a reversed light/dark cycle^{8,10}. Slices were prepared during the subjective day to prevent phase shifts. The delays between time of preparation and time of recording were similar for day and night. Day cells were recorded between circadian time (CT) 4 and 8 (refs 2, 10) and night cells between CT 13 and 20, with CT 0 corresponding to light onset. The day–night differences described here are unlikely to be due to after-effects of light, because rats from both the day and night group were exposed for at least 3 h to light before preparation. Strictly speaking, however, we cannot fully rule out effects of prolonged, repeated light exposures and therefore prefer to use the term ‘diurnal’ instead of ‘circadian’ modulation. Following national guidelines on animal experiments, rats were anaesthetized^{8,17} and transcardially perfused with 30–40 ml of ice-cold artificial cerebrospinal fluid (ACSF; composition (in mM): 124.0 NaCl, 3.5 KCl, 1.0 NaH_2PO_4 , 26.2 NaHCO_3 , 1.3 MgSO_4 , 2.5 CaCl_2 , 10.0 D(+)-glucose and 0.0125 bicuculline methiodide; gassed with a mixture of 95% O_2 and 5% CO_2 ; pH 7.4). Transverse slices (200 μm) of hypothalamus were cut with a vibroslicer (Campden Instruments). Flow rate in the recording chamber (32 °C) was 1.5–2.5 ml min^{-1} .

Current clamp recordings

Current clamp recordings were conducted using the perforated patch technique¹⁰. The pipette solution contained (in mM): 135.0 potassium gluconate, 10.0 KCl, 10.0 HEPES buffer, 0.5 EGTA (pH adjusted to 7.3 with KOH; osmolality 275–285 mOsm kg^{-1}). Gramicidin and amphotericin B were dissolved in dimethylsulphoxide and added to this solution (final concentrations 250 and 5 $\mu\text{g ml}^{-1}$, respectively). We found that the combination of these substances produced a lower series resistance (28–67 M Ω) than either substance alone, without loss of recording stability. Patch pipettes (4–7 M Ω) were positioned close to SCN neurons under constant positive pressure, using an upright fixed-stage microscope (Axioskop, Zeiss) equipped with a 40X water-immersion lens (numerical aperture 0.75) with Hoffman modulation contrast. After gigaseal (>3 G Ω) formation, perforation was regularly monitored by measuring capacitive current transients evoked by a –20-mV voltage step. Most recordings were made from the dorsomedial SCN, where cluster I neurons are predominant¹⁷. Voltage traces were acquired at a sampling rate of 20 kHz using an Axopatch-1D or Axoclamp-2B amplifier and analysed with pClamp (version 6.0.4, Axon Instruments). Membrane potential values were determined from the time-averaged readout of the electrode amplifier and corrected for the liquid junction potential (–12 mV). Our criterion for labelling a cell as ‘oscillating’ was an oscillation peak–valley amplitude of at least 3 mV and a frequency of at least 0.5 Hz. A Hamming window was used to compute power spectra.

Voltage clamp recordings

Ca^{2+} currents in SCN slices were studied in whole-cell mode because it allowed a lower series resistance (11–32 M Ω ; 80% compensation) than perforated patch mode. Owing to their very high input resistance, SCN neurons in slices are electrotonically compact and allow for voltage clamping of Ca^{2+} current. The pipette medium contained (in mM): 115.0 caesium gluconate, 20.0 tetraethylammonium chloride, 10.0 HEPES buffer, 0.5 EGTA, 2.0 MgCl_2 , 20.0 sodium phosphocreatine, 2.0 Na_2ATP , 0.3 Na_3GTP and 0.1 leupeptin (pH 7.3; osmolality 275–285 mOsm kg^{-1} ; pipette resistance 4–7 M Ω). The ACSF contained (in mM): 68.0 NaCl, 3.5 KCl, 1.0 NaH_2PO_4 , 26.2 NaHCO_3 , 1.3 MgSO_4 , 2.5 CaCl_2 , 10.0 D(+)-glucose, 60 tetraethylammonium chloride, 3.0 CsCl, 5.0 4-aminopyridine, 0.0125 bicuculline and 0.001 tetrodotoxin (pH 7.4). Ca^{2+} instead of Ba^{2+} was used as charge carrier because Ca^{2+} -dependent inactivation of Ca^{2+} current may be subject to diurnal modulation and is thus interesting to take into account. Currents were filtered at 1–2 kHz and digitized every 100–200 μs using an Axopatch-1D amplifier. Corrections were made for the liquid junction potential (–9 mV). Cells were accepted for analysis when two criteria indicative of good voltage clamp control were satisfied: the current–voltage relationship showed a graded increment in Ca^{2+} current in the range –60 to –20 mV; and the onset of Ca^{2+} current was not delayed with respect to the voltage step. We used a P/4 protocol to subtract leak currents.

Drugs and statistics

All drugs were administered by bath application. Stock solutions were prepared by dissolving nimodipine (RBI) in 100% ethanol and TTX (Alomone Labs), ω -conotoxin GVIA and ω -agatoxin IVB (Peptides International) in distilled water. These solutions were diluted in ACSF to their final concentrations. Nimodipine was protected from light during these procedures. When using ω -agatoxin IVB, the perfusion system was pre-exposed to a 0.01% cytochrome c solution to prevent the peptide binding to tubing and containers. The toxin was applied for at least 25 min. NiCl_2 was obtained from Sigma. Groups of cells were statistically compared with the use of Mann–Whitney’s *U*-test and Wilcoxon’s matched-

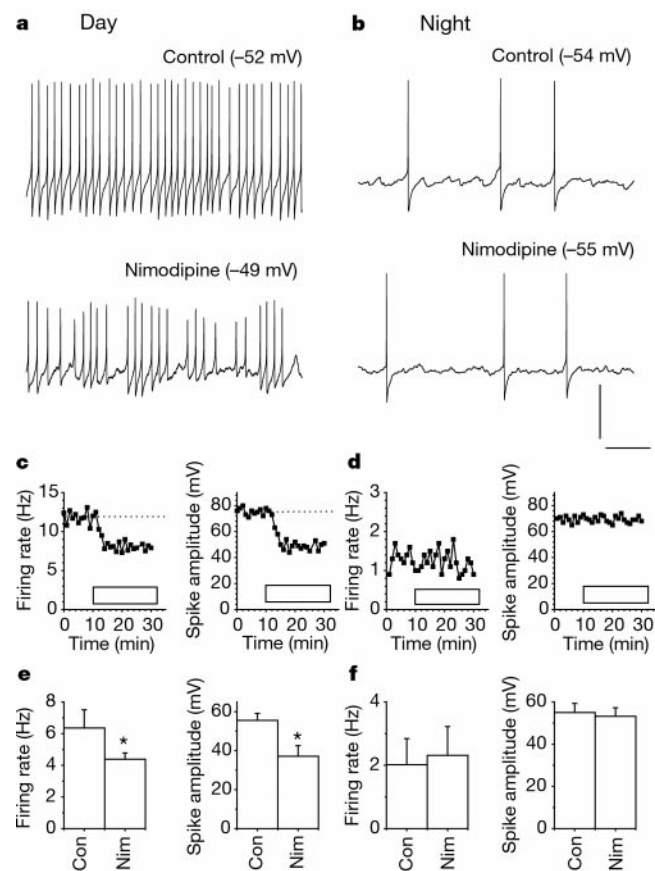


Figure 4 Differential effect of nimodipine on spontaneous firing of SCN neurons during the day and night. **a**, Day: nimodipine (2 μM) caused a reduction in spike amplitude, amplitude of the spike afterhyperpolarization and spontaneous firing rate. **b**, Night: nimodipine failed to affect these parameters. Calibration bars (**a**, **b**), 40 mV, 500 ms. **c**, **d**, Time course of nimodipine effects on firing rate and spike amplitude for day (**c**) and night (**d**). Open bar, period of nimodipine application. **e**, **f**, Decrease in firing rate and spike amplitude was significant across the population of day cells (**e**; asterisk, $P < 0.05$) but absent at night (**f**). Amplitude of the spike afterhyperpolarization in day cells decreased from 19 ± 1 mV (control) to 12 ± 1 mV (nimodipine) ($P < 0.05$).

pairs signed-rank test for unpaired and paired comparisons, respectively. Fisher's exact test was used to evaluate the difference in prevalence of oscillations across populations of day and night cells. Correlation coefficients between firing rates and oscillation frequencies were determined by computing Spearman's rank correlation. All numerical values both in text and graphs represent the mean $\pm$ s.e.m.

Received 6 November 2001; accepted 15 February 2002.

Published online 3 March 2002, DOI 10.1038/nature728.

- Inouye, S.-I. T. & Kawamura, H. Persistence of circadian rhythmicity in a mammalian hypothalamic "island" containing the suprachiasmatic nucleus. *Proc. Natl Acad. Sci. USA* **76**, 5962–5966 (1979).
- Green, D. J. & Gillette, R. Circadian rhythm of firing rate recorded from single cells in the rat suprachiasmatic brain slice. *Brain Res.* **245**, 198–200 (1982).
- Schwartz, W. J., Gross, R. A. & Morton, M. T. The suprachiasmatic nuclei contain a tetrodotoxin-resistant circadian pacemaker. *Proc. Natl Acad. Sci. USA* **84**, 1694–1698 (1987).
- Welsh, D. K., Logothetis, D. E., Meister, M. & Reppert, S. M. Individual neurons dissociated from rat suprachiasmatic nucleus express independently phased circadian firing rhythms. *Neuron* **14**, 697–706 (1995).
- Reppert, S. M. & Weaver, D. R. Molecular analysis of mammalian circadian rhythms. *Annu. Rev. Physiol.* **63**, 547–676 (2001).
- Allada, R., Emery, P., Takahashi, J. S. & Rosbash, M. Stopping time: the genetics of fly and mouse circadian clocks. *Annu. Rev. Neurosci.* **24**, 1091–1119 (2001).
- Huang, R.-C. Sodium and calcium currents in acutely dissociated neurons from rat suprachiasmatic nucleus. *J. Neurophysiol.* **70**, 1692–1703 (1993).
- Pennartz, C. M. A., Bierlaagh, M. A. & Geurtsen, A. M. S. Cellular mechanisms underlying spontaneous firing in rat suprachiasmatic nucleus: involvement of a slowly inactivating component of sodium current. *J. Neurophysiol.* **78**, 1811–1825 (1997).
- Jiang, Z.-G., Yang, Y., Liu, Z.-P. & Allen, C. N. Membrane properties and synaptic inputs of suprachiasmatic nucleus neurons in rat brain slices. *J. Physiol. (Lond.)* **499**, 141–159 (1997).
- De Jeu, M. T. G., Hermes, M. H. L. J. & Pennartz, C. M. A. Circadian modulation of membrane properties in slices of rat suprachiasmatic nucleus. *Neuroreport* **9**, 3725–3729 (1998).
- Mogul, D. J. & Fox, A. P. Evidence for multiple types of Ca^{2+} channels in acutely isolated hippocampal CA3 neurones of the guinea-pig. *J. Physiol. (Lond.)* **433**, 259–281 (1991).
- Avery, R. B. & Johnston, D. Multiple channel types contribute to the low-voltage-activated calcium current in hippocampal CA3 pyramidal neurons. *J. Neurosci.* **16**, 5567–5582 (1996).
- Randall, A. & Tsien, R. W. Pharmacological dissection of multiple types of Ca^{2+} channel currents in rat cerebellar granule neurons. *J. Neurosci.* **15**, 2995–3012 (1995).
- Fox, A. P., Nowycky, M. C. & Tsien, R. W. Kinetic and pharmacological properties distinguishing three types of calcium currents in chick sensory neurons. *J. Physiol. (Lond.)* **394**, 149–172 (1987).
- Blaxter, T. J., Carlen, P. L. & Niesen, C. Pharmacological and anatomical separation of calcium currents in rat dentate granule neurons *in vitro*. *J. Physiol. (Lond.)* **412**, 93–112 (1989).
- Adams, M. E., Mintz, I. M., Reilly, M. D., Thanabal, V. & Bean, B. P. Structure and properties of

- ω -Agatoxin IVB, a new antagonist of P-type calcium channels. *Mol. Pharmacol.* **44**, 681–688 (1993).
- Pennartz, C. M. A., Geurtsen, A. M. S., De Jeu, M. T. G., Sluiter, A. A. & Hermes, M. H. L. J. Electrophysiological and morphological heterogeneity of neurons in slices of rat suprachiasmatic nucleus. *J. Physiol. (Lond.)* **506**, 775–793 (1998).
- Michel, S., Geusz, M. E., Zaritsky, J. J. & Block, G. D. Circadian rhythm in membrane conductance expressed in isolated neurons. *Science* **259**, 239–241 (1993).
- D'Souza, T. & Dryer, S. E. A cationic channel regulated by a vertebrate intrinsic circadian oscillator. *Nature* **382**, 165–167 (1996).
- Takahashi, J. S. & Zatz, M. Regulation of circadian rhythmicity. *Science* **217**, 1104–1111 (1982).
- Nedergaard, S., Flatman, J. A. & Engberg, I. Nifedipine- and ω -conotoxin-sensitive Ca^{2+} conductances in guinea-pig substantia nigra pars compacta neurones. *J. Physiol. (Lond.)* **466**, 727–747 (1993).
- Kasai, H. & Neher, E. Dihydropyridine-sensitive and ω -conotoxin-sensitive calcium channels in a mammalian neuroblastoma-glioma cell line. *J. Physiol. (Lond.)* **448**, 161–188 (1992).
- Fisher, T. E. & Bourque, C. W. Calcium-channel subtypes in the somata and axon terminals of magnocellular neurosecretory cells. *Trends Neurosci.* **19**, 440–444 (1996).
- Colwell, C. S. Circadian modulation of calcium levels in cells in the suprachiasmatic nucleus. *Eur. J. Neurosci.* **12**, 571–576 (2000).
- Block, G. D., Khalsa, S. B. S., McMahon, D. G., Michel, S. & Geusz, M. Biological clocks in the retina: Cellular mechanisms of biological timekeeping. *Int. Rev. Cytol.* **146**, 83–143 (1993).
- Murphy, T. H., Worley, P. F. & Baraban, J. M. L-type voltage-sensitive calcium channels mediate synaptic activation of immediate early genes. *Neuron* **7**, 625–635 (1991).
- Dolmetsch, R. E., Xu, K. & Lewis, R. S. Calcium oscillations increase the efficiency and specificity of gene expression. *Nature* **392**, 933–936 (1998).
- Catterall, W. A. Structure and regulation of voltage-gated Ca^{2+} channels. *Annu. Rev. Cell Dev. Biol.* **16**, 521–555 (2000).
- Ko, G. Y.-P., Ko, M. L. & Dryer, S. E. Circadian regulation of cGMP-gated cationic channels of chick retinal cones: Erk MAP kinase and Ca^{2+} /calmodulin-dependent protein kinase II. *Neuron* **29**, 255–266 (2001).
- O'Rourke, B., Backx, P. H. & Marban, E. Phosphorylation-independent modulation of L-type calcium channels by magnesium–nucleotide complexes. *Science* **257**, 245–248 (1992).

Acknowledgements

We thank G. Borst, J. Meijer, T. Mulder, S. Taverna and M. Verhage for their comments on the manuscript and F. Lopes da Silva for his stimulating advice. This work was supported by the Netherlands Organization for Scientific Research (NWO).

Correspondence and requests for materials should be addressed to C.M.A.P. (e-mail: c.pennartz@nih.knaw.nl).

Killing activity of neutrophils is mediated through activation of proteases by K^+ flux

Emer P. Reeves^{*†}, Hui Lu^{*†}, Hugues Lortat Jacobs[‡], Carlo G. M. Messina^{*}, Steve Bolsover[§], Giorgio Gabella^{||}, Eric O. Potma[¶], Alice Warley[#], Jürgen Roes[✉] & Anthony W. Segal^{*}

^{*}Centre for Molecular Medicine, Departments of § Physiology and || Anatomy and [✉] Windeyer Institute of Medical Sciences, University College London, 5 University Street, London WC1E 6JJ, UK

[‡] Institut de Biologie Structurale, 41 rue Horowitz, 38027 Grenoble, France

[¶] Ultrafast Laser and Spectroscopy Laboratory, Materials Science Centre, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

[#] Rayne Institute, St Thomas' Hospital, Lambeth Palace Road, London SE1 7EH, UK

[†] These authors contributed equally to this work.

According to the hitherto accepted view, neutrophils kill ingested microorganisms by subjecting them to high concentrations of highly toxic reactive oxygen species (ROS) and bringing about myeloperoxidase-catalysed halogenation. We show here that this simple scheme, which for many years has served as a satisfactory working hypothesis, is inadequate. We find that mice made deficient in neutrophil-granule proteases but normal in respect of superoxide production and iodinating capacity, are unable to resist staphylococcal and candidal infections. We also show that activation provokes the influx of an enormous concentration of ROS into the endocytic vacuole. The resulting accumulation of anionic charge is compensated for by a surge of K^+ ions that cross the membrane in a pH-dependent manner. The consequent rise in ionic strength engenders the release of cationic granule proteins, including elastase and cathepsin G, from the anionic sulphated proteoglycan matrix. We show that it is the proteases, thus activated, that are primarily responsible for the destruction of the bacteria.

Metchnikoff discovered that 'phagocytes' confer immunity by engulfing invading microbes¹. It was supposed that killing was effected by the contents of the cytoplasmic granules released into the phagocytic vacuoles in which the microbes were encapsulated. This hypothesis was supplanted^{2,3} by the supposition that the killing agents were ROS^{4,5}, supported by the discovery of chronic granulomatous disease (CGD), a human condition characterized by profound susceptibility to bacterial and fungal infection⁶. The finding that phagocytes from these individuals are unable to generate ROS⁶ or to kill microbes efficiently was taken to imply that the respiratory burst promotes killing by generating toxic superoxide^{4,5} and H_2O_2 (ref. 7). The H_2O_2 was thought also to exert an indirect effect, as substrate for myeloperoxidase (MPO)-catalysed halogenation⁷.

According to this prevailing ROS model, a lack of proteases should not affect the killing ability of neutrophils. Although earlier studies had shown a requirement for neutrophil elastase to kill some Gram-negative bacilli⁸, and for both cathepsin G and elastase to protect against infection by *Aspergillus fumigatus*⁹, we extended these observations to microbes—*Staphylococcus aureus*¹⁰, the commonest cause of infection in CGD⁶, and *Candida albicans*—that, it was thought, could be killed only with the intervention of oxygen-dependent processes, because mice lacking MPO are abnormally sensitive to this pathogen¹¹.

Infections in protease-deficient mice

Figure 1 shows the results of intravenous injections of *S. aureus* (Fig. 1a) or *C. albicans* (Fig. 1b) on the survival of protease-deficient mice and *S. aureus* on CGD mice. Both organisms were much more virulent in the mice lacking both cathepsin G and elastase. A striking feature was the differential sensitivity of the two species of microbe to the two proteases. Cathepsin-G-deficient mice resisted *C. albicans* but not *S. aureus*, whereas the reverse was true for those lacking elastase (see also Supplementary Information).

The microbicidal activity of purified neutrophils determined *in vitro* (Fig. 1c, d) was the mirror image of the susceptibility *in vivo* (Fig. 1a, b). Under the same experimental conditions—microbial iodination—a composite measure of phagocytosis,

degranulation, oxidase activity and MPO activity was normal in these mice (Fig. 1f).

The addition of a cocktail of protease inhibitors to a suspension of human cells and bacteria also severely impaired killing (Fig. 1e); as expected, killing was also greatly reduced in CGD cells or those treated with the oxidase inhibitor diphenylene iodonium (DPI)¹².

Conditions within the phagocytic vacuole

ROS and protease deficiencies lead to comparable reductions in killing efficiency, implying that they act together on the internalized microbe. To clarify the nature of their interaction we set out to define conditions prevailing in the vacuole immediately after phagocytosis, when killing occurs¹³.

To estimate vacuolar volume, cross-sectional areas of phagocytic vacuoles were determined from electron micrographs, which indicated that the volume of the vacuole during the respiratory burst is about $0.2 \mu m^3$ (see Fig. 4a, b). For each molecule of O_2 consumed 4 superoxide (O_2^-) ions are generated (Fig. 2f). Between 0.2 (ref. 14) and 0.5 fmol (this study) of oxygen is consumed for each bacterium engulfed, resulting in an intravacuolar O_2^- release of about $4 mol l^{-1}$. The steady-state concentrations of O_2^- and H_2O_2 attained are uncertain but have been surmised to be in the micromolar range¹⁵.

The large amounts of O_2^- , and therefore of its products, generated in the vacuole lead to the consideration that it might exert electrostatic or other physical effects, rather than merely providing the substrate for MPO. This conjecture was strengthened by observations on the cells of patients with incomplete CGD, or 'variant CGD'¹⁶, which still display impaired microbicidal activity despite generating about $0.5 mol l^{-1}$ O_2^- in the vacuole. We accordingly examined phagocytic vacuoles by electron microscopy to search for evidence of oxidase-dependent changes in morphology. The vacuoles of normal cells were strikingly different in appearance from those of a patient with X-linked CGD (X-CGD), and another suffering from variant X-CGD with 12% oxidase activity (Fig. 2a–c). The granule contents were uniformly dispersed throughout the normal vacuoles but remained clumped in most of the CGD vacuoles and in those treated with DPI (not shown).

Knowing the vacuolar volume we determined the concentration of granule protein that entered vacuoles (Fig. 2d) as being about 0.5 g ml^{-1} . The pH in the vacuole rises from about 6.0 to 7.8–8.0 soon after phagocytosis, in consequence of oxidase activity^{13,17}, despite the entry of acid granule contents¹⁸. The O_2^- is released into the vacuole together with the granule contents, which exert a pH-buffering effect. This buffer capacity was estimated from the titre of KOH required to raise the pH from 5.5, at which the granules are maintained in the intact cell¹⁸, to a final value of 8.0 (Fig. 2e)¹³. This proved to be about $400 \mu\text{mol}$ KOH per gram of granule protein.

Oxidase pumps K^+ into the vacuole

We show here that the oxidase disperses the granules by pumping large amounts of K^+ into the vacuole.

The passage of electrons across the vacuolar membrane is

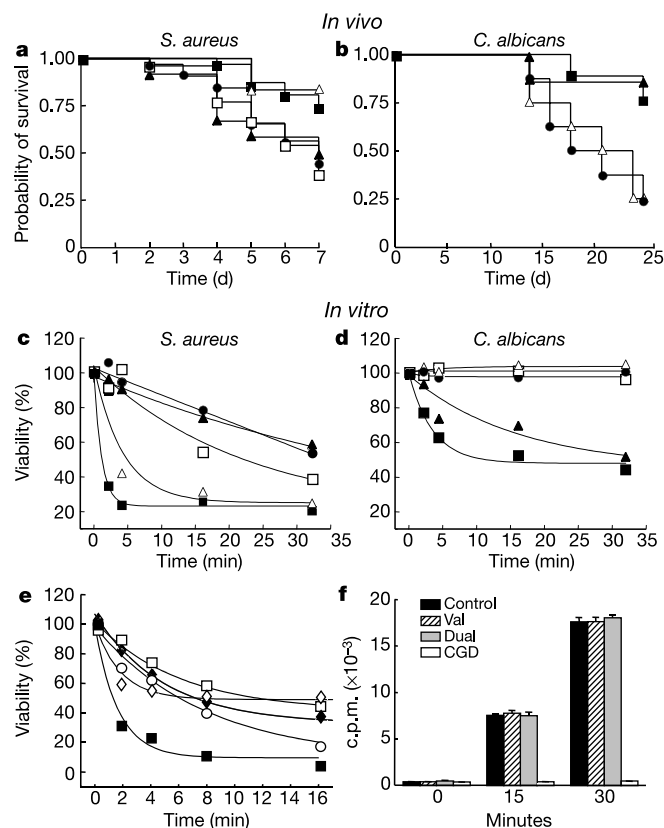


Figure 1 Susceptibility of protease-deficient mice to bacterial and fungal infection. **a**, **b**, Survival probability plots (Kaplan–Meier) of wild-type mice (filled squares, $n=32$), or mice deficient in neutrophil elastase⁹ (open triangles, $n=12$), cathepsin G⁹ (filled triangles, $n=12$), elastase and cathepsin G⁹ (filled circles, $n=31$) or p47^{phox} (ref. 9) (open squares, $n=24$), injected intravenously with 4×10^7 *S. aureus* cells (a) or 10^4 *C. albicans* cells (b) (filled squares, $n=9$; open triangles, $n=8$; filled triangles, $n=7$; filled circles, $n=8$). The rates of decreased probability of survival for all mice were significantly different ($P < 0.05$) from wild type except for elastase-deficient (a) and cathepsin-G-deficient (b) animals, for which $P = 0.1$. **c**, **d**, Killing of *S. aureus* (c) and *C. albicans* (d) by mouse neutrophils (symbols as in a) *in vitro*. Killing rates for all mice were significantly different from wild type ($P < 0.01$) except for elastase-deficient (c) and cathepsin-G-deficient (d) cells ($P = 0.4$ and 0.9 , respectively). **e**, Killing of *S. aureus* by human neutrophils *in vitro* to show effects of addition of valinomycin (open circles), DPI (open diamonds) and protease inhibitors (filled diamonds). All compounds inhibited killing to the level seen with CGD cells (open squares) when compared with control cells (filled squares) ($P < 0.005$) and had no effect on bacteria. **f**, Time course of iodination by neutrophils from wild-type mice, mice deficient in cathepsin G and elastase (Dual) and mice deficient in p47^{phox} (CGD) exposed to *S. aureus*. The effect of the addition of valinomycin to normal cells is also shown (Val).

electrogenic¹⁹. The required charge compensation is commonly thought to be encompassed solely by the transfer of protons into the vacuole^{19–21} (Fig. 2f). Compensation of an electron by a proton is pH-neutral and cannot explain the observed elevation in vacuolar pH (ref. 13) that occurs despite the entry of acidic granule proteins into the vacuole. Cl^- ions could migrate from the vacuole to the cytosol, but the concentrations of 150 mM taken into the vacuole from the extracellular medium would not suffice to bring about

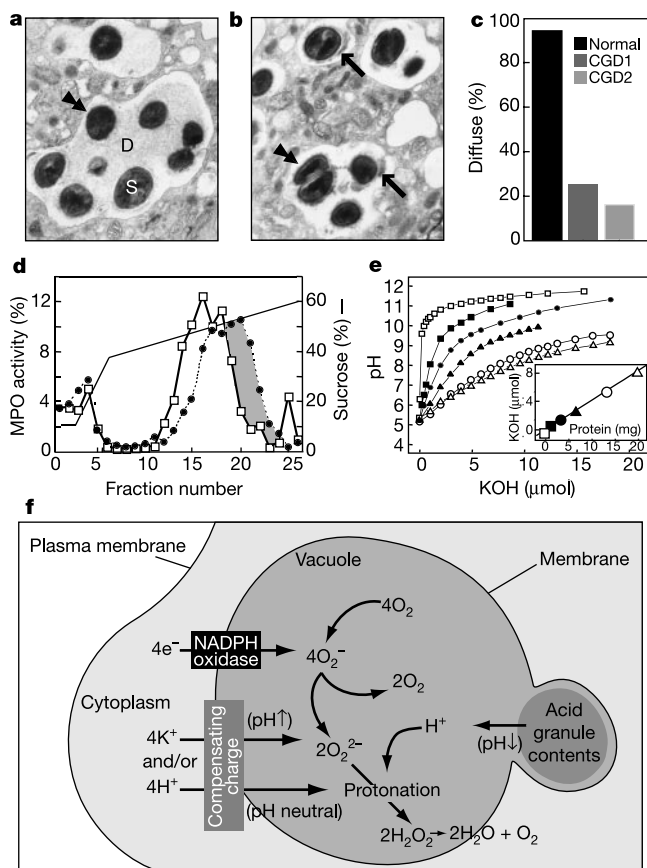


Figure 2 Appearance, concentration and buffering capacity of granule contents in the phagocytic vacuole. **a**, **b**, Granule contents within the vacuole (a) were classified as diffuse, because of their homogenous distribution (D), or clumped (arrows in b). Bacteria are indicated by double arrowheads. **c**, The proportion of vacuoles with a diffuse appearance in normal neutrophils ($n=280$) and those from patients with X-linked CGD with complete absence of oxidase activity (CGD1, $n=244$) or a variant patient with 12% oxidase activity (CGD2, $n=206$) ($P = 10^{-148}$ and 10^{-68} between normal and X-CGD and between normal and variant X-CGD patients, respectively, by χ^2 analysis). **d**, Movement of granular proteins to the vacuole during phagocytosis of *S. aureus*. Distribution of MPO in phagocytosing (filled circles) compared with resting (open squares) neutrophils. Shading indicates movement of MPO to dense vacuoles containing bacteria. MPO composed 26% of the total granule protein. The protein content of 10^7 cells was 1.2 mg , of which $9.4 \pm 1.0\%$ (mean $\pm$ s.e.m., $n=4$) was in the granule fraction; 24%, or $\sim 0.1 \text{ pg}$ per vacuole, migrated to the vacuoles after the phagocytosis of 25 bacteria per cell. **e**, Buffering capacity of granule proteins (1 ml) at various concentrations titrated against KOH. Inset, amount of KOH required to elevate the pH from 5.5 to 8.0 for different amounts of protein. **f**, Schematic representation of factors involved in compensation of charge across the phagosomal membrane. The pH in the vacuole rises¹³, despite the influx of acidic granule contents, because of the consumption of H^+ within this compartment by the protonation of O_2^- and O_2^{2-} . If all the charge produced by pumping electrons across the phagosomal membrane were compensated by H^+ ions from the cytosol, the pH in the vacuole would not rise. We show here that part of the charge is compensated for by K^+ , which elevates the pH in a regulated manner and causes the vacuole to become hypertonic.

charge compensation. Cl^- has, in any case, been shown to move in the opposite direction in stimulated cells²². The rectifying movement of cations, in particular K^+ , from the cytoplasm where its concentration is high, into the vacuole, provides an alternative mechanism. First we studied the effect of valinomycin, a specific K^+ ionophore, on O_2^- production and O_2 consumption, and also that of 4-aminopyridine, a nonspecific K^+ -channel blocker²³ (Fig. 3a); both processes were markedly accelerated by the former and inhibited (albeit to a smaller extent) by the latter.

Because there is no known indicator of the required specificity and sensitivity for the measurement of K^+ at the high levels predicted in the vacuole, we determined its release into the medium after stimulation with the PKC agonist, 12-*O*-tetradecanoylphorbol-13-acetate (TPA), a potent activator of the oxidase. In response to this agent, which induces the oxidase to secrete O_2^- across the plasma membrane to the exterior, K^+ was released

into the medium, and this was blocked by DPI (Fig. 3b). Measurements with the K^+ electrode were only semiquantitative, so $^{86}\text{Rb}^+$, commonly used as a marker of K^+ in transport studies²⁴, was used for quantification. Secretion of this ion was suppressed by DPI (Fig. 3c) and is thus contingent upon O_2^- generation rather than on PKC activation by the phorbol ester. To show that this release was spatially coupled to the O_2^- -generating machinery and not to general leakage from respiring cells, we measured secretion in conditions favouring the flux of K^+ , Rb^+ or O_2^- into the closed phagocytic vacuole²⁵ encapsulating an engulfed bacterium. In these circumstances, release to the exterior was not observed, even though an equivalent amount of oxygen was consumed.

We used electron-probe X-ray microanalysis to obtain a direct estimate of the K^+ content of the phagosome (Fig. 3e). In DPI-treated cells the vacuolar K^+ concentration was much less than that in untreated control cells (mean $\pm$ s.e.m. $122 \pm 15 \text{ mmol g}^{-1}$ ($n = 35$) and $314 \pm 22 \text{ mmol g}^{-1}$ ($n = 40$), respectively; $P < 0.001$ by Wilcoxon rank-sum), clearly implicating the oxidase in the accumulation of K^+ in this compartment. A wide range of values is measured in normal vacuoles in consequence of the asynchronous nature of phagocytosis. To take into account the wide spread of results obtained with the microprobe technique, and to detect a skewing of the distribution towards higher concentrations, we compared the upper quartiles of K^+ measurements in the vacuoles with those in the corresponding cytosol. We found them to be significantly higher in the control cells ($P < 0.03$ by Wilcoxon rank-sum), no different in the valinomycin-treated cells ($P = 0.16$) and much lower in cells exposed to DPI ($P = 0.0003$). Microprobe analysis gave a median concentration of K^+ of 290 mmol g^{-1} in the cytosol of normal cells, which we take to correspond to the experimentally measured concentration of 125 mM (ref. 27). This ratio leads to a value of 270 mM for the upper quartile in normal vacuoles, corresponding to the microprobe measurement of 631 mmol g^{-1} . Thus we can conclude that K^+ concentrations in the range 200–300 mM are attained in the vacuole (see also Supplementary Information).

What sets the level of charge compensation by K^+ ? Each molecule of O_2^- generates one OH^- ion. If all the charge were to be compensated for by K^+ the OH^- would overwhelm the buffering capacity of the granules and the pH would be excessively elevated. The regulation of the proportion of compensating charge made up by K^+ seems to depend on the pH itself. The efflux of $^{86}\text{Rb}^+$ to the exterior after stimulation with TPA was reduced when the pH of the medium was elevated to 8.0, and almost eliminated at pH 8.5, even though O_2^- production was virtually unchanged at the higher pH (Fig. 3d). At high pH the charge is likely to be compensated for almost entirely by H^+ . In this way the amount of K^+ entering the vacuole could be linked to degranulation. The more granule contents are released into the vacuole, the more the pH is depressed and the greater the buffering capacity. A parallel increase in the proportion of the charge compensated for by K^+ would be required for transport of this ion to be impeded by a rise in pH.

K^+ influx renders the vacuole hypertonic

After a lag of about 1 min after stimulation, 10^7 neutrophils produce approximately 300 nmol of O_2^- in the ensuing 2 min (Fig. 3a, b). With a cellular volume of $345 \mu\text{m}^3$ and a K^+ concentration of 125 mM (ref. 26), 2.9×10^9 cells occupy 1 ml and contain $125 \mu\text{mol}$ of K^+ , of which, judging from the $^{86}\text{Rb}^+$ measurement, roughly 4.5%, or $5.6 \mu\text{mol}$, would be released. Over the same interval the cells would produce about $90 \mu\text{mol}$ of O_2^- , a molar ratio to K^+ secretion of approximately 16:1. The release of 4 mol l^{-1} O_2^- in the vacuole then leads to the expected intravacuolar K^+ concentration in the range 200–300 mM, very similar to that measured by X-ray microanalysis. One electron charge compensated by K^+ generates one OH^- , so the levels of OH^- that would be produced agree satisfactorily with the alkali titre required to raise

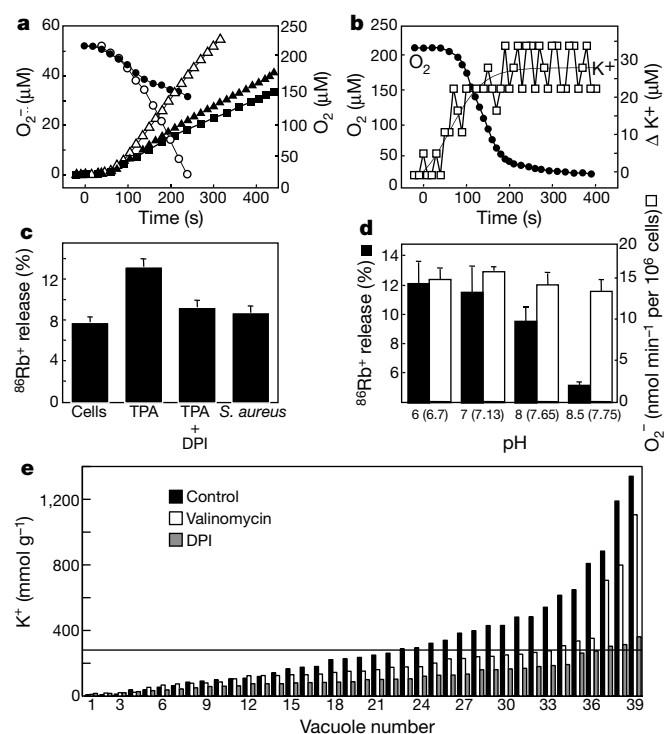


Figure 3 K^+ and ^{86}Rb transport associated with the generation of O_2^- in neutrophils. **a**, The rate of O_2^- release from 10^6 ml^{-1} cells is stimulated by valinomycin ($3 \mu\text{M}$, 3 min; open triangles) and inhibited by 4-aminopyridine (4 mM , 3 min; filled squares) compared with control cells (filled triangles) (representative of three experiments) in response to TPA (10 ng ml^{-1}). O_2 consumption by 10^7 cells phagocytosing *S. aureus* in the absence (filled circles) and the presence (open circles) of valinomycin is also shown. **b**, Time course of release of K^+ into medium (open squares) in relation to oxygen consumption (filled circles) after stimulation of 4×10^7 cells ml^{-1} with TPA ($1 \mu\text{g ml}^{-1}$). **c**, $^{86}\text{Rb}^+$ release from 10^7 cells before (Cells) or after stimulation with TPA ($1 \mu\text{g ml}^{-1}$), or phagocytosis of *S. aureus*. Oxygen consumption was equivalent after these stimuli and was inhibited by DPI. **d**, $^{86}\text{Rb}^+$ release (black columns), but not O_2^- generation (white columns), was inhibited by elevating the pH to 8 or more. The pH values of the external buffer and of cytosol (in parentheses) are shown. (All results are means $\pm$ s.e.m. from at least three measurements). **e**, K^+ concentrations in phagocytic vacuoles of neutrophils after the phagocytosis of *S. aureus*, and the effects on this of the addition of valinomycin and DPI, as determined by electron probe X-ray microanalysis. The horizontal line depicts the mean concentration (280 mmol g^{-1}) in control cytosol. (Median, mean, n and s.e.m. of values in vacuoles were: 236, 322, 39 and 50; 152, 213, 39 and 34; and 100, 121, 39 and 13 mmol g^{-1} for cells treated with control, valinomycin and DPI, respectively. The corresponding values in cytosols were 291, 280, 41 and 22; 211, 269, 46 and 38; and 278, 345, 37 and 46.)

the pH in the presence of the predicted concentration of granule contents.

Tonicity in the vacuole must initially be high. The K^+ driven into the vacuole by the burst is added to the 150 mM NaCl, incorporated from the extracellular medium when the vacuole forms. Dilution of these salts occurs only later, as swelling of the vacuoles¹³ (Fig. 4a, b) sets in well after the end of the respiratory burst. To understand why swelling does not ensue immediately, we first used coherent anti-Stokes Raman scattering (CARS) laser-scanning microscopy²⁷ to determine whether the vacuolar membrane was unusually impermeable to water. The results showed the membrane surrounding the vacuole to be permeable to H_2O (Fig. 4c–e), with a permeability constant of $6 \mu m s^{-1}$ or higher, roughly similar to that of the plasma membrane of *Dictyostelium discoideum* cells²⁷; thus another explanation is required for the delay in swelling. It turns out that the vacuole is constrained by a surrounding meshwork of cytoskeletal proteins, as seen in macrophages²⁸, including vinculin and paxillin. These proteins initially accumulate around the vacuole and then slowly disperse as swelling proceeds (Fig. 4f, g). When the membrane-permeant fungal toxin jasplakinolide²⁹ was used to stabilize filamentous actin, swelling of the vacuoles was prevented (Fig. 4b).

Because valinomycin has such a profound effect in suppressing bacterial killing and vacuolar swelling, despite occasioning a marked enhancement of the respiratory burst, we measured its effect on vacuolar pH. We found that valinomycin caused a large decrease in vacuolar pH (Fig. 5e), to the low values observed in cells lacking the oxidase activity, either because of CGD¹³ or inhibition by DPI¹². A possible explanation for this effect is that almost twice as much oxygen is consumed for each bacterium engulfed in valinomycin-treated cells, and thus more cations are required for charge compensation. K^+ ions are lost from the vacuole through the valinomycin channel, so that the charge must be compensated for by the passage of protons, with consequent acidification. The pH in

the vacuoles of the valinomycin-treated cells is below the optimum for both elastase and cathepsin G (Fig. 5f), a contributing factor to the impairments of killing (Fig. 1e) and digestion (Fig. 5d) induced by the ionophore.

With regard to the force within the vacuole responsible for swelling, the osmotic effect of the added K^+ is in itself probably insufficient. Fairly high levels of K^+ were measured in the presence of valinomycin, yet there was no swelling; protease inhibitors, which should not perturb K^+ fluxes, also prevented swelling (Fig. 4b). Swelling seems instead to be linked to the digestion of vacuolar contents and thus probably results from the osmotic pressure exerted by the products of this process. Digestion and swelling are both impaired in the absence of the oxidase, either naturally in CGD¹³ or after inhibition by DPI, by valinomycin or by protease inhibitors (Fig. 5d), or when microbes are replaced by protease-resistant latex particles¹³. The osmotic forces in the vacuole must be considerable as they cause the ingested bacteria to shrink to about half their original volume (Fig. 4a inset, b) before partly recovering as the vacuoles swell.

Hypertonicity activates granule proteases

How then might the K^+ enter into the killing process? Clearly it can do so by enabling the pH in the vacuole to rise as it substitutes for H^+ in compensating for the potential difference across the membrane generated by the oxidase. However, it also exerts a quite separate effect by activating the granule enzymes. The granules contain a strongly anionic sulphated proteoglycan matrix to which the cationic proteases are tightly bound³⁰. A model for this is the affinity of some of these proteases for heparin as shown by surface plasmon resonance in Fig. 5a. In this bound state they cannot reach the ingested microbe and must be solubilized to become functional. We propose that this activation is brought about by the hypertonic K^+ , driven into the vacuole by the NADPH oxidase. Figure 5b shows that most of the granule protein is solubilized at the K^+ concentrations

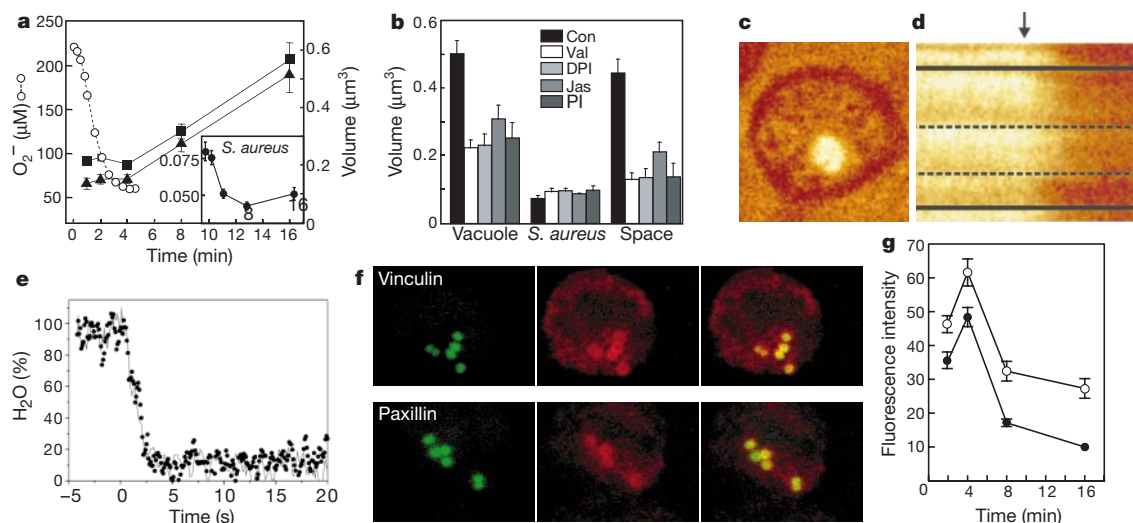


Figure 4 Vacuolar swelling and its control. **a**, Time course of the changes in the volume of the phagocytic vacuole (filled squares), the enclosed bacterium (filled circles, insert) and the intervening space (filled triangles) determined by electron microscopy on 24, 43, 65, 45 and 24 vacuoles measured after 1, 2, 4, 8 and 16 min respectively. The bacteria re-swelled significantly between 8 and 16 min ($P < 0.05$ by Student's t -test). The time course of oxygen consumption is also shown (open circles). **b**, The effect on the size of the vacuoles of exposing cells to valinomycin (Val, 3 μM , $n = 106$), DPI (5 μM , $n = 44$), jasplakinolide (1 μM , $n = 74$) and protease inhibitors (PI, $n = 30$) compared with untreated cells (Con, $n = 111$) at 16 min after phagocytosis. All additions resulted in highly significant differences ($P < 0.001$) from control cells. **c**, CARS image of neutrophil with superimposed fluorescence image of ANTS-labelled, phagocytosed *C. albicans*. **d**, Real-

time permeability experiment showing the CARS signal from water on the left and the ensuing decrease in signal after rapid exchange (arrow) of aqueous medium against isotonic 2H_2O buffer. Solid lines mark the plasma membrane of the cell and dashed lines indicate the edges of the intracellular vacuole. **e**, Dynamics of water efflux from the vacuole upon rapid $H_2O/^2H_2O$ exchange. Dots refer to experimental data points. The solid line fitted to the data reveals an efflux time of H_2O from the vacuole of ~ 0.1 s. **f**, Fluorescein-labelled *S. aureus* (approximate diameter 0.8 μm) in phagocytic vacuoles (green, left panels) of neutrophils stained for vinculin or paxillin (red, middle panels). These cytoskeletal proteins surround the vacuole (composite images, right panels). **g**, Amounts of vinculin (open circles) and paxillin (filled circles) around the vacuole diminish with time as swelling occurs (each point is the mean of 50 measurements).

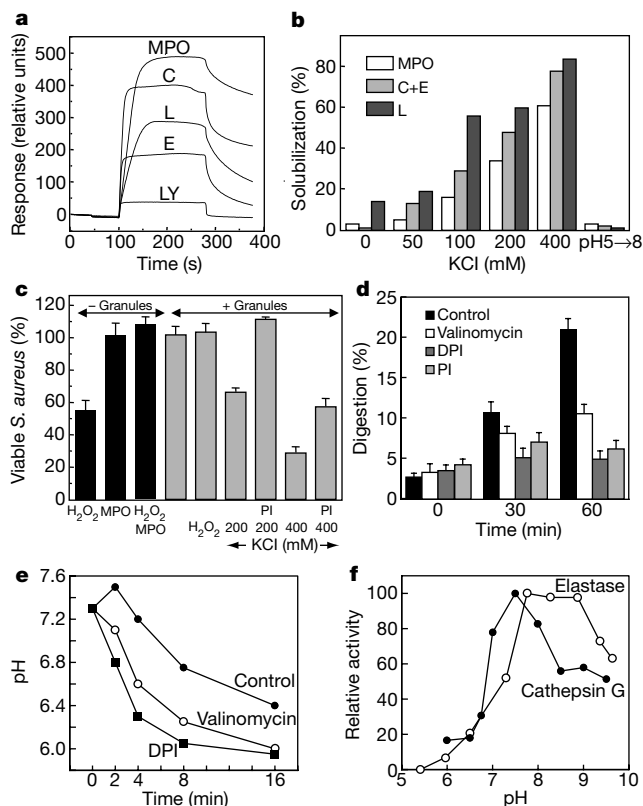


Figure 5 Activation of the microbicidal proteases requires K^+ and neutral pH. **a**, The cationic granule proteins bind strongly to proteoglycans. Surface plasmon resonance plots of binding of pure MPO (K_d 20.4 nM), cathepsin G (C, K_d 5.3 nM), lactoferrin (L, K_d 28.3 nM) and elastase (E, K_d 8.9 nM) to heparin. Lysozyme (LY) did not bind significantly. All proteins were present at $4 \mu\text{g ml}^{-1}$. **b**, MPO, cathepsin G and elastase (C+E) and lactoferrin (L) were released from the granule matrix at elevated salt concentrations but not by increasing the pH from 5 to 8 (granules 20 mg ml^{-1}). **c**, The effect of H_2O_2 (100 mM), alone or with MPO (5 mg ml^{-1}), and purified neutrophil granules (44 mg ml^{-1}), in the presence and absence of KCl and/or protease inhibitors (PI) on the viability of *S. aureus*. All results are means $\pm$ s.e.m. from at least three measurements. **d**, Effect of valinomycin, DPI and protease inhibitors (PI) on digestion of ^{35}S -labelled, killed, *S. aureus*. **e**, Effect of DPI and valinomycin on vacuolar pH over time. Each point is the mean of at least three measurements with fluorescein-coated bacteria as pH indicators^{13,17}. pH values at 2 min were significantly different ($P < 0.05$ by Student's *t*-test). **f**, Relation between pH and activity of cathepsin G and elastase. Each point is the mean of three measurements.

achieved in the vacuole. Elevated pH on its own was without effect. Isolated granules were not bactericidal but became so at higher KCl concentrations, an effect that was partly abrogated by protease inhibitors (Fig. 5c). The fact that some of the antimicrobial activity of the cationic proteins is independent of their enzymic function³¹ could account for this inhibitor-resistant killing. The concentrations of H_2O_2 developed in the vacuole are not known but it seems unlikely that they could exceed 100 mM, because at such high levels the catalase activity of MPO supervenes^{32,33}; the action of H_2O_2 on microbes was therefore measured at this concentration. Only limited microbicidal activity was recorded, and even this was suppressed completely by the addition of granules. MPO also inhibited killing, despite the presence in the medium of Cl^- , a substrate for chlorination reactions. Lower concentrations of H_2O_2 , down to 100 μM , also failed to kill, with or without MPO (not shown).

Discussion

Our results have uncovered a previously unsuspected mechanism of antimicrobial activity in the phagocyte. In brief, the O_2^- -generating

system causes an influx of K^+ into the phagocytic vacuole with an attendant rise in pH to the optimal level for the activity of the granule proteases. Moreover, these become active on release from the anionic proteoglycan matrix at the elevated ionic strength.

It is essential that the volume of the vacuole be restricted for the requisite hypertonicity to develop. This requires the integrity of the cytoskeletal network, and disruption of this by microbial products could offer a mechanism of virulence by inhibiting the activation of the granule proteins³⁴.

The identity and nature of the K^+ channel will be of interest; it might or might not be the well-studied H^+ channel. One suggestion has been that gp91^{phox}, the flavocytochrome *b* of the NADPH oxidase, is itself the channel^{35,36}, although contradictory evidence exists^{37,38}.

One might question the need for such an elaborate activation system. A possible explanation lies in the very large numbers of neutrophils that infiltrate sites of acute inflammation and the potential of their enzymes to damage autologous tissues if released from cells in a freely soluble, active, form. For example, the mouse lacking both cathepsin G and elastase is relatively resistant to endotoxin shock, in which damage to the endothelium by the endotoxin is followed by neutrophil attack; this implicates these two proteases in the tissue damage that normally occurs in this situation⁹. The hazard of injurious effects from these enzymes in normal tissues is reduced by packaging them in an inactive form adsorbed on acidic proteoglycans, from which they are released and activated only by a combination of the unusual conditions of hypertonicity and alkalinity prevailing in the phagocytic vacuole.

Our demonstration that ROS generation and MPO activity are not themselves sufficient to kill key model target organisms is important, not only because of the insight it affords into normal immunity, but also because of the light it throws on pathological mechanisms. Early theories implicating oxygen radicals in tissue damage³⁹ stemmed from the apparent toxicity of these agents against microbes, which are much tougher than human cells. Some of this early supportive evidence will need to be reassessed in the light of the above developments^{4,5}. Experiments with MPO were generally performed at what has been shown here to be unphysiologically low concentrations of enzyme and H_2O_2 and at too low a pH (ref. 7).

MPO still seems to be important in the killing process^{7,11}. It might help to protect the microbicidal enzymes against oxidative damage. In preliminary studies we have found cathepsin G to be very sensitive to oxidation by H_2O_2 and to be inactivated at a greatly increased rate in phagocytosing neutrophils treated with azide to inhibit MPO. Similar mechanisms of oxidative inactivation of degradative enzymes could explain the accelerated deposition of atheromatous material observed in MPO-deficient mice^{40,41}. □

Methods

Infection studies

Staphylococcus aureus (NCTC 12981) cells (4×10^7) or *C. albicans* cells (10^6) from an overnight culture were injected into the tail veins of mice deficient in cathepsin G and/or neutrophil elastase, CGD mice deficient in p47^{phox} or normal mice, all on the strain 129 genetic background⁹. Survival was monitored every 12 h for 7 d (*S. aureus*) and 24 d (*C. albicans*), respectively.

Measurements of neutrophil function *in vitro*

Neutrophils were purified from human blood¹³ and suspended in PBS (140 mM NaCl, 10 mM KCl, 10 mM NaH_2PO_4 , 5 mM glucose, pH 7.3). All mixing of microbes with neutrophils *in vitro* was done in the rapidly stirred chamber of an oxygen electrode at 37 °C. All additions of chemicals to cell suspensions were made without subsequent incubation unless stated otherwise.

Thioglycollate-elicited mouse neutrophils (2.8×10^7) (ref. 9) in 0.5 ml were mixed with 5×10^6 *S. aureus* or 2.5×10^6 *C. albicans* cells. Killing was measured as described¹³, omitting lysostaphin. All killing studies *in vitro* were performed in duplicate with cells from three to five mice, and colonies were counted on three aliquots taken at each time point.

Iodination by mouse neutrophils was measured on cells purified from thioglycollate (3%)–induced peritoneal exudates and with the use of immune mouse serum as opsonin⁴². Results of three experiments with three measurements in each are shown in Fig. 1f.

To measure the effect of valinomycin (3 μM), DPI (5 μM) or protease inhibitors (10 $\mu\text{g ml}^{-1}$ leupeptin, *N*-tosyl-L-leucine chloromethyl ketone (TLCK), pepstatin A, aprotinin and diisopropylfluorophosphate 1 mM) on killing by human neutrophils, 10^8 cells in 1 ml PBS were mixed with 10^5 IgG-opsonized *S. aureus*, in the presence and in the absence of valinomycin or protease inhibitors, and then treated as above. To study the effect of jasplakinolide (1 μM) without inhibiting phagocytosis, cells were mixed with bacteria for 2 min; the agent was then added and cells were incubated at 37 °C for a further 14 min.

O_2^- generation was determined by superoxide-dismutase-inhibitable reduction of cytochrome *c*⁴ and oxygen consumption with an oxygen electrode¹⁴. Digestion of heat-killed (65 °C, 30 min) ³⁵S-methionine-labelled *S. aureus* cells was expressed as the percentage of trichloroacetic-acid (10%) -soluble radioactivity released into the supernatant¹³.

Survival analysis

Survival of mice was analysed by Kaplan–Meier survival analysis. Killing *in vitro* was analysed by determining the rate of loss of viability of the organisms by regression analysis of a logarithmic transformation of the numbers of surviving microbes in relation to time. Slopes were compared with a two-tailed *t*-test.

Electron and confocal microscopy

To assess the sizes of phagocytic vacuoles and their contents, 5×10^7 fresh human neutrophils were mixed with 5×10^8 IgG-opsonized *S. aureus* cells in a rapidly stirred chamber at 37 °C. At timed intervals, aliquots were taken into a fixative consisting of glutaraldehyde (1.5%), picric acid (0.1%) in sodium cacodylate buffer (0.1 M, pH 7.2). They were postfixed in osmium tetroxide, dehydrated in ethanol and embedded in Araldite. Sections were collected on copper grids, contrast-stained with uranyl acetate and lead citrate and viewed at a magnification of $\times 16,000$. Morphometry was performed by projecting the photographic electron-microscopy negatives through a microfilm reader, giving a further enlargement of $\times 9$ or $\times 12$. The relevant structures (and the calibration bars) were traced on paper and the profiles were then measured with a digitizing tablet. Only the volumes of phagocytic vacuoles that were judged to have been sectioned in the region of maximum width were measured.

For assessment of the dispersion of granule contents in the vacuole, the cells were treated as above and the incubation was ended at 4 min. Granule contents in the vacuoles were visually assessed as dispersed or not by two independent 'blinded' observers.

Immunohistochemical analysis of phagocytosing neutrophils was conducted as described previously⁴³ with antibodies against paxillin and vinculin (from Santa Cruz). The fluorescence intensity around 50 vacuoles was measured at each time point and compared with that in the surrounding cytosol by using a Leica TCS NT confocal microscope.

Measurement of K^+ and $^{86}\text{Rb}^+$ fluxes

K^+ release from neutrophils ($4 \times 10^7 \text{ ml}^{-1}$) stimulated with $1 \mu\text{g ml}^{-1}$ TPA was measured continuously with a K^+ electrode (Thermo Russell, Model 93-3199) at 37 °C. The cells were kept in PBS and the buffer was exchanged to a K^+ -free PBS just before each measurement. K^+ efflux was measured in the absence and in the presence of 5 μM DPI. K^+ concentration was calculated from a standard curve.

For $^{86}\text{Rb}^+$ efflux, 10^7 neutrophils were suspended in 1 ml 136 mM NaCl, 1 mM MgSO_4 , 1 mM CaCl_2 , 5 mM KCl, 5.6 mM glucose, 20 mM HEPES pH 7.4 (HEPES buffer), 2.5 $\mu\text{Ci } ^{86}\text{Rb}^+$ and 0.1 mg ml^{-1} bovine serum albumin (BSA) and incubated at 30 °C for 30 min, washed three times in the buffer without the $^{86}\text{Rb}^+$ and then placed in a stirred oxygenated chamber at 37 °C. Cells were stimulated with either 2×10^8 IgG-opsonized *S. aureus* or $1 \mu\text{g ml}^{-1}$ TPA for 3.5 min. Stimulated cells were immediately cooled to 4 °C and centrifuged through Ficoll/Hypaque at 400g for 10 min at 4 °C. Radioactivity in the supernatant above the Ficoll/Hypaque and in the neutrophil pellet was removed and counted in a liquid-scintillation counter. $^{86}\text{Rb}^+$ efflux was also measured in 20 mM HEPES buffer at pH 6, 7, 8 and 8.5.

K^+ concentrations inside the phagocytic vacuole and cytoplasm were measured by X-ray microanalysis on snap-frozen, freeze-dried cells. Neutrophils (with and without DPI or valinomycin) were mixed with IgG-opsonized *S. aureus* at 37 °C for 3 min and centrifuged at 800g for 10 s; drops of the thick cell pellet were collected on a piece of Millipore filter paper and cryofixed by being plunged into liquid propane. The cryofixed pellets were stored under liquid nitrogen. Cryosections 250 nm thick were cut at -120°C in an RMC MTXL cryo-ultramicrotome. The sections were supported on Pioloform-covered nickel grids (100 mesh hexagonal) and freeze-dried overnight under temperature-controlled conditions. The sections were coated with a thin layer of carbon in the vacuum chamber and analysed with a Zeiss EM10 electron microscope fitted with an Oxford Instruments detector and PGT Avalon 8000 microanalysis system. Areas within the vacuole space and in the adjacent cytoplasm were analysed with a reduced-area raster for 60 s of live time at an accelerating voltage of 80 kV and an beam current of 1 nA in STEM mode at ambient temperature. Spectra were processed and quantification was achieved with the peak-to-continuum ratio method⁴⁴. Results are expressed as mmol g^{-1} dry weight.

Granule and protein purification, enzyme activity and plasmon resonance

Human buffy coat neutrophils were disrupted by nitrogen cavitation in Break buffer (10 mM KCl, 3 mM NaCl, 4 mM MgCl_2 , 10 mM PIPES pH 7.2) containing protease inhibitors (leupeptin, TLCK and pepstatin A, each at $1 \mu\text{g ml}^{-1}$). Post-nuclear supernatants (PNSs) were prepared by centrifugation at 400g at 4 °C for 10 min.

The PNS was centrifuged for 1 h at 100,000g in a Beckman SW41 head at 4 °C on discontinuous gradients of 33% sucrose on 55% sucrose (w/w), and the granules were

removed from the interface. Human MPO was purified from granules to a Reinheit Zahl (Rz) value (A_{430}/A_{280}) of 0.82 (ref. 45).

For purification of cathepsin G, neutrophil elastase, lactoferrin and lysozyme, the granules were homogenized in Break buffer and 1% cetyltrimethylammonium bromide. The extract was centrifuged at 25,000g for 10 min. The supernatant was loaded on a Fast Flow S-Sepharose column (1.5 cm $\times$ 20 cm) in 10 mM phosphate buffer pH 6.95 and eluted with a linear gradient of NaCl (10 mM to 1 M). Lysozyme, cathepsin G and lactoferrin were further purified on a Mono S column (1 ml) with the same buffer and salt gradient. Elastase was further purified on heparin–agarose in the same buffer and salt gradient followed by gel filtration on a Superdex 75 column (all resins from Pharmacia). All proteins were at least 95% pure and identities were confirmed by matrix-assisted laser desorption/ionization mass spectrometry. Protein binding to heparin was measured by surface plasmon resonance (Biacore) as described⁴⁶.

To determine the quantity of granule proteins in the vacuole, bacteria were labelled with [³⁵S]methionine, heat-killed (70 °C for 10 min), opsonized with human IgG and mixed with cells at a ratio of 20:1 at 37 °C for 4 min. Unopsonized bacteria were used as a control. Phagocytosis was stopped by placing cells into ice-cold PBS in which they were washed twice and then lysed by nitrogen cavitation in Break buffer containing 11.2% (w/w) sucrose. The PNS was centrifuged into a 12-ml continuous sucrose gradient from 30% to 60% (w/w) at 150,000g for 2 h in a Beckman SW41 head. Fractions (0.5 ml) were counted for bacteria and assayed for MPO (with *o*-dianisidine) or protein (with Bradford reagent from Bio-Rad) with BSA as standard.

Membranes were removed from granules⁴⁷ by centrifugation through Percoll. To measure the buffering capacity of granule proteins, various concentrations of granules lacking membranes, suspended in normal saline, were titrated against KOH and the pH was measured.

The pH dependence of the activities of elastase and cathepsin G⁴⁸ (both from Sigma) was measured exactly as described.

Killing of *S. aureus* by granule proteins

Granules without membranes were prepared in the absence of protease inhibitors. Bacteria were incubated for 6 min with H_2O_2 (100 mM) in the presence or absence of MPO (5 mg ml^{-1}) or granules (44 mg ml^{-1}) in PBS at 37 °C. Granules were also incubated in the presence of KCl (200 or 400 mM), with or without protease inhibitors as above.

Measurement of vacuolar permeability to water

Purified neutrophils were left to adhere to poly-(L-lysine)-coated slides for 15 min at 37 °C. The slides were washed with PBS and then overlaid for 15 min with a suspension of *C. albicans* labelled with the fluorophore 8-aminonaphthalene-1,3,6-trisulphonic acid (ANTS; Molecular Probes) and opsonized with human IgG. Unattached *Candida* cells were washed away and nonlinear coherent CARS microscopy was performed, in combination with isotopic exchange, as described⁴⁷.

Vacuolar and cytosolic pH

Vacuolar pH was determined from the excitation spectrum of fluorescein-labelled, IgG-coated *S. aureus*¹³ exactly as described¹⁷. Cytosolic pH was measured in cells in suspension with the fluorescent indicator 2',7'-bis-(carboxyethyl)-5(6)-carboxyfluorescein⁴⁹, and pH values were standardized with phosphate-buffered or HEPES-buffered salt solutions; cells were equilibrated in the presence of either *p*-(trifluoromethoxy)phenylhydrazine (10 μM) or nigericin (10 μM).

Received 12 December 2001; accepted 23 January 2002.

1. Metchnikoff, B. *Immunity in Infective Diseases* (Cambridge Univ. Press, 1905).
2. Sbarra, A. J. & Karnovsky, M. L. The biochemical basis of phagocytosis. 1. Metabolic changes during the ingestion of particles by polymorphonuclear leukocytes. *J. Biol. Chem.* **234**, 1355–1362 (1959).
3. Mandell, G. L. Bactericidal activity of aerobic and anaerobic polymorphonuclear neutrophils. *Infect. Immun.* **9**, 337–341 (1974).
4. Babior, B. M., Curnutte, J. T. & Kipnes, R. S. Biological defense mechanisms. Evidence for the participation of superoxide in bacterial killing by xanthine oxidase. *J. Lab. Clin. Med.* **85**, 235–244 (1975).
5. Babior, B. M., Kipnes, R. S. & Curnutte, J. T. Biological defence mechanisms: the production by leukocytes of superoxide, a potential bactericidal agent. *J. Clin. Invest.* **52**, 741–744 (1973).
6. Thrasher, A. J., Keep, N. H., Wientjes, F. & Segal, A. W. Chronic granulomatous disease. *Biochim. Biophys. Acta* **1227**, 1–24 (1994).
7. Klebanoff, S. J. Antimicrobial mechanisms in neutrophilic polymorphonuclear leukocytes. *Semin. Hematol.* **12**, 117–142 (1975).
8. Belaouaj, A. *et al.* Mice lacking neutrophil elastase reveal impaired host defense against Gram-negative bacterial sepsis. *Nature Med.* **4**, 615–618 (1998).
9. Tkalec, J. *et al.* Impaired immunity and enhanced resistance to endotoxin in the absence of neutrophil elastase and cathepsin G. *Immunity* **12**, 201–210 (2000).
10. Hampton, M. B., Kettle, A. J. & Winterbourn, C. C. Involvement of superoxide and myeloperoxidase in oxygen-dependent killing of *Staphylococcus aureus* by neutrophils. *Infect. Immun.* **64**, 3512–3517 (1996).
11. Aratani, Y., Koyama, H., Nyui, S., Suzuki, K., Kura, F. & Maeda, N. Severe impairment in early host defense against *Candida albicans* in mice deficient in myeloperoxidase. *Infect. Immun.* **67**, 1828–1836 (1999).
12. Cross, A. R. & Jones, O. T. The effect of the inhibitor diphenylene iodonium on the superoxide-generating system of neutrophils. Specific labelling of a component polypeptide of the oxidase. *Biochem. J.* **237**, 111–116 (1986).
13. Segal, A. W., Geisow, M., Garcia, R., Harper, A. & Miller, R. The respiratory burst of phagocytic cells is associated with a rise in vacuolar pH. *Nature* **290**, 406–409 (1981).

14. Segal, A. W. & Coade, S. B. Kinetics of oxygen consumption by phagocytosing human neutrophils. *Biochem. Biophys. Res. Commun.* **84**, 611–617 (1978).
15. Hampton, M. B., Kettle, A. J. & Winterbourn, C. C. Inside the neutrophil phagosome: oxidants, myeloperoxidase, and bacterial killing. *Blood* **92**, 3007–3017 (1998).
16. Bu-Ghanim, H. N., Segal, A. W., Keep, N. H. & Casimir, C. M. Molecular analysis in three cases of X91-variant chronic granulomatous disease. *Blood* **86**, 3575–3582 (1995).
17. Jiang, Q., Griffin, D. A., Barofsky, D. F. & Hurst, J. K. Intraphagosomal chlorination dynamics and yields determined using unique fluorescent bacterial mimics. *Chem. Res. Toxicol.* **10**, 1080–1089 (1997).
18. Styrts, B. & Klempner, M. S. Internal pH of human neutrophil lysosomes. *FEBS Lett.* **149**, 113–116 (1982).
19. Henderson, L. M., Chappell, J. B. & Jones, O. T. The superoxide-generating NADPH oxidase of human neutrophils is electrogenic and associated with an H⁺ channel. *Biochem. J.* **246**, 325–329 (1987).
20. Nanda, A. & Grinstein, S. Protein kinase C activates an H⁺ (equivalent) conductance in the plasma membrane of human neutrophils. *Proc. Natl Acad. Sci. USA* **88**, 10816–10820 (1991).
21. DeCoursey, T. E., Cherny, V. V., Zhou, W. & Thomas, L. L. Simultaneous activation of NADPH oxidase-related proton and electron currents in human neutrophils. *Proc. Natl Acad. Sci. USA* **97**, 6885–6889 (2000).
22. Menegazzi, R., Busetto, S., Dri, P., Cramer, R. & Patriarca, P. Chloride ion efflux regulates adherence, spreading, and respiratory burst of neutrophils stimulated by tumor necrosis factor- α (TNF) on biologic surfaces. *J. Cell Biol.* **135**, 511–522 (1996).
23. Clapp, L. H. & Tinker, A. Potassium channels in the vasculature. *Curr. Opin. Nephrol. Hypertens.* **7**, 91–98 (1998).
24. Love, W. D. & Burch, G. E. A comparison of potassium⁴², rubidium⁸⁶, and cesium¹³⁴ as tracers of potassium in the study of cation metabolism of human erythrocytes *in vitro*. *J. Lab. Clin. Med.* **41**, 351–362 (1953).
25. Segal, A. W. & Meshulam, T. Production of superoxide by neutrophils: a reappraisal. *FEBS Lett.* **100**, 27–32 (1979).
26. Ince, C. *et al.* Intracellular K⁺, Na⁺ and Cl⁻ concentrations and membrane potential in human monocytes. *Biochim. Biophys. Acta* **905**, 195–204 (1987).
27. Potma, E., de Boei, W. P., van Haastert, P. J. & Wiersma, D. A. Real-time visualization hydrodynamics in single living cells. *Proc. Natl Acad. Sci. USA* **98**, 1577–1582 (2001).
28. Aderem, A. & Underhill, D. M. Mechanisms of phagocytosis in macrophages. *Annu. Rev. Immunol.* **17**, 593–623 (1999).
29. Rizoli, S. B., Rotstein, O. D., Parodo, J., Phillips, M. J. & Kapus, A. Hypertonic inhibition of exocytosis in neutrophils: central role for osmotic actin skeleton remodeling. *Am. J. Physiol. Cell Physiol.* **279**, C619–C633 (2000).
30. Kolset, S. O. & Gallagher, J. T. Proteoglycans in haemopoietic cells. *Biochim. Biophys. Acta* **1032**, 191–211 (1990).
31. Odeberg, H. & Olsson, I. Antibacterial activity of cationic proteins from human granulocytes. *J. Clin. Invest.* **56**, 1118–1124 (1975).
32. Winterbourn, C. C., Garcia, R. C. & Segal, A. W. Production of the superoxide adduct of myeloperoxidase (compound III) by stimulated human neutrophils and its reactivity with hydrogen peroxide and chloride. *Biochem. J.* **228**, 583–592 (1985).
33. Kettle, A. J. & Winterbourn, C. C. A kinetic analysis of the catalase activity of myeloperoxidase. *Biochemistry* **40**, 10204–10212 (2001).
34. Guerin, I. & de Chastellier, C. Pathogenic mycobacteria disrupt the macrophage actin filament network. *Infect. Immun.* **68**, 2655–2662 (2000).
35. Henderson, L. M. & Meech, R. W. Evidence that the product of the human X-linked CGD gene, gp91-phox, is a voltage-gated H⁺ pathway. *J. Gen. Physiol.* **114**, 771–786 (1999).
36. Maturana, A. *et al.* Heme histidine ligands within gp91^{phox} modulate proton conduction by the phagocyte NADPH oxidase. *J. Biol. Chem.* **276**, 30277–30284 (2001).
37. Nanda, A., Romanek, R., Curnutte, J. T. & Grinstein, S. Assessment of the contribution of the cytochrome *b* moiety of the NADPH oxidase to the transmembrane H⁺ conductance of leukocytes. *J. Biol. Chem.* **269**, 27280–27285 (1994).
38. De Coursey, T. E., Cherny, V. V., Morgan, D., Katz, B. Z. & Dinauer, M. C. The gp91^{phox} component of NADPH oxidase is not the voltage-gated proton channel in phagocytes, but it helps. *J. Biol. Chem.* **276**, 36063–36066 (2001).
39. McCord, J. M. & Wong, K. *in Oxygen Free Radicals and Tissue Damage* (ed. Fitzsimons, D. W.) 343–360 (Excerpta Medica, Amsterdam, 1979).
40. Klebanoff, S. J. & Pincus, S. H. Hydrogen peroxide utilization in myeloperoxidase-deficient leukocytes: a possible microbicidal control mechanism. *J. Clin. Invest.* **50**, 2226–2229 (1971).
41. Brennan, M. L. *et al.* Increased atherosclerosis in myeloperoxidase-deficient mice. *J. Clin. Invest.* **107**, 419–430 (2001).
42. Klebanoff, S. J. & Clark, R. A. Iodination by human polymorphonuclear leukocytes: a re-evaluation. *J. Lab. Clin. Med.* **89**, 675–686 (1977).
43. Grogan, A. *et al.* Cytosolic phox proteins interact with and regulate the assembly of coronin in neutrophils. *J. Cell Sci.* **110**, 3071–3081 (1997).
44. Hall, T. A. & Gupta, B. L. *in Principles of Analytical Electron Microscopy* (eds Joy, D. C., Romig, A. D. & Goldstein, J. I.) 219–248 (Plenum, London, 1986).
45. Olsen, R. L. & Little, C. Purification and some properties of myeloperoxidase and eosinophil peroxidase from human blood. *Biochem. J.* **209**, 781–787 (1983).
46. Sadir, R., Baleux, F., Grosdidier, A., Imbert, A. & Lortat-Jacob, H. Characterization of the stromal cell-derived factor-1- α -heparin complex. *J. Biol. Chem.* **276**, 8288–8296 (2001).
47. Vita, F. *et al.* Preparation of membrane fractions from human neutrophil granules: A simple method. *Methods Cell Sci.* **19**, 197–205 (1997).
48. Barrett, A. J. Cathepsin G. *Methods Enzymol.* **80C**, 561–565 (1981).
49. Amos, B. J., Pocock, G. & Richards, C. D. On the role of bicarbonate as a hydrogen ion buffer in rat CNS neurones. *Exp. Physiol.* **81**, 623–632 (1996).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank W. Gratzner, P. Rich, F. Ashcroft, M. Brand, R. Vaughan-Jones and M. Duchon for discussions; E. Hawe for statistical analysis; A. Hankin and E. C. Davis for technical assistance; and A. Scott for illustration and art work. The Wellcome Trust and Chronic Granulomatous Disease Research Trust provided financial support.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.W.S. (e-mail: rmhaase@ucl.ac.uk).

A topographically forced asymmetry in the martian circulation and climate

Mark I. Richardson* & R. John Wilson†

* Division of Geological and Planetary Sciences, California Institute of Technology, MC 150-21, Pasadena, California 91125, USA

† Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, PO Box 308, Princeton, New Jersey 08542, USA

Large seasonal and hemispheric asymmetries in the martian climate system are generally ascribed to variations in solar heating associated with orbital eccentricity¹. As the orbital elements slowly change (over a period of $>10^4$ years), characteristics of the climate such as dustiness and the vigour of atmospheric circulation are thought to vary^{2–5}, as should asymmetries in the climate (for example, the deposition of water ice at the northern versus the southern pole). Such orbitally driven climate change might be responsible for the observed layering in Mars' polar deposits by modulating deposition of dust and water ice^{3,5,6}. Most current theories assume that climate asymmetries completely reverse as the angular distance between equinox and perihelion changes by 180° . Here we describe a major climate mechanism that will not precess in this way. We show that Mars' global north–south elevation difference forces a dominant southern summer Hadley circulation that is independent of perihelion timing. The Hadley circulation, a tropical overturning cell responsible for trade winds, largely controls interhemispheric transport of water and the bulk dustiness of the atmosphere^{7–11}. The topography therefore imprints a strong handedness on climate, with water ice and the active formation of polar layered deposits more likely in the north.

Models suggest that the meridional circulation on Mars is characterized by a dominant cross-equatorial ('solstitial') Hadley circulation except during a brief transitional period near the equinoxes^{7,8,12,13} (Fig. 1). The circulation has upwelling in the summer tropics, resulting in adiabatic cooling, and downwelling in the winter tropics, which produces adiabatic warming (on Earth, these upwelling and downwelling belts create the cloud and rain characteristic of the equatorial rain forests, and the dry, clear atmosphere over the subtropical deserts). In this way, the dominant Hadley cell transports heat from the summer hemisphere into the winter hemisphere. It is accompanied by a very much weaker cell operating exclusively in the summer hemisphere. The martian mean-meridional circulation is similar to that of the Earth^{14–16}, although it is more exaggerated owing to the lack of martian oceans to buffer seasonal oscillations in surface temperature. On Mars, however, the solstitial cells are not of equal intensity or extent—the dominant southern summer Hadley cell is roughly twice as intense as that of the northern summer^{7,12}. This is partly due to the planetary eccentricity (0.093) and the fact that perihelion occurs just before the southern summer solstice, resulting in solar incident radiation being as much as 44% higher than experienced in northern summer—southern summer is shorter, but more intense than northern summer.

The influence of the Hadley circulation on climate, through the transport and mixing of dust and water vapour, can be illustrated with output from the Geophysical Fluid Dynamics Laboratory (GFDL) Mars general circulation model (GCM)^{8,11} (Fig. 1). These simulations show that the vertical distribution of dust within the atmosphere is largely determined by the depth of the atmosphere mixed by the Hadley cell. The simulations also illustrate the fact that interhemispheric transport of water is dominated by conveyance in the upper-level Hadley circulation (Fig. 1b). The low-level return flow brings relatively dry air into the summer hemisphere. The

vertical distribution of dust plays an important role in determining the bulk dustiness of the atmosphere^{8,9}, while the nature of the water cycle and the stability of surface water deposits in communication with the atmosphere is critically sensitive to the vigour of interhemispheric water transport^{10,11}. For these reasons, the Hadley circulation is a major factor in the martian climate and is the focus of this study.

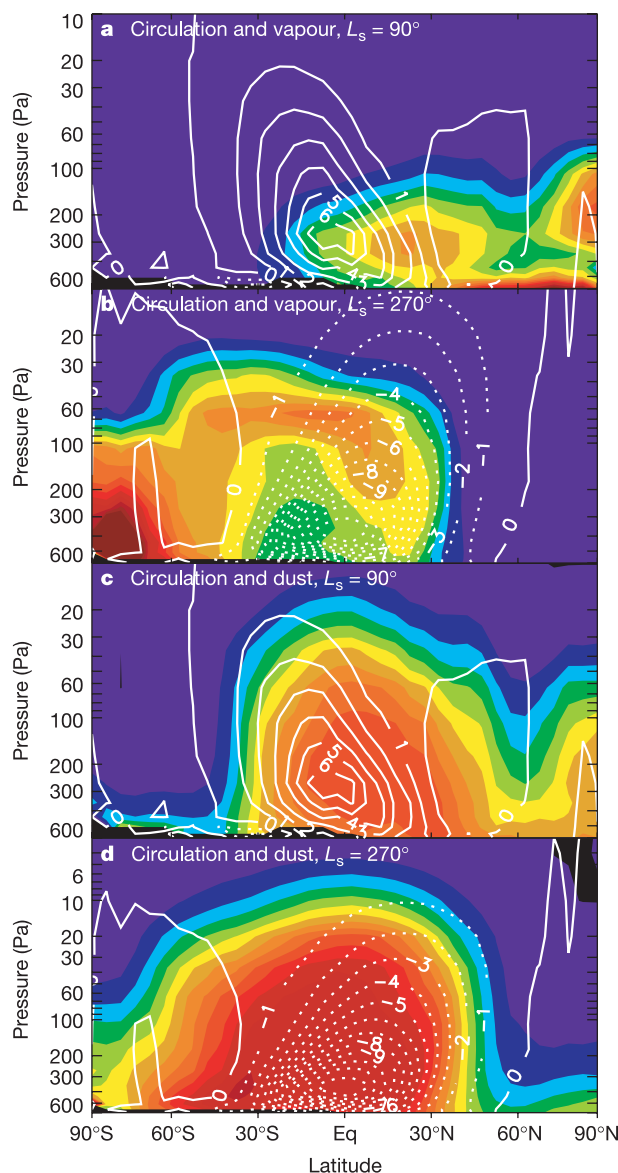


Figure 1 Mars GCM output illustrating the Hadley circulation's strong control of water vapour transport and the global mixing of dust. Zonal-average output are shown for northern summer, $L_s = 90^\circ$, and southern summer, $L_s = 270^\circ$, where L_s is the angular measure of the season, measured in degrees from 0° at northern spring equinox. Stream-function contours are linear in units of 10^8 kg s^{-1} . Panels **a** and **b** show water vapour mass mixing ratio; **c** and **d** show a quantity linearly related to dust mass mixing ratio. Water is supplied only from the summer polar cap, and subsequent evolution depends on transport and atmospheric condensation. Dust injection at the surface is based on stability. Dust and condensed water are subject to gravitational settling. The dust distribution (a significant atmospheric opacity source) is used to calculate radiative heating rates. Water vapour is assumed to be radiatively inactive. The stream function is the conventional eulerian mean constructed from zonal-average mass convergence. Transformed eulerian mean provides a better representation of transport of trace species when eddy heat and momentum transports are large. For the martian tropical atmosphere, the two formulations are essentially identical (compare the vapour distribution and stream function in southern summer).

In the Earth's atmosphere, the seasonally asymmetric Hadley circulations aggregate to yield two roughly equal cells in the annual average (albeit with cells significantly stronger than would be generated by application of annual-average heating¹⁶). Our simulation of the martian meridional circulation demonstrates that this is not true for Mars (Fig. 2a). The annual average of the mean-meridional (mass transport stream function) circulation shows a clear bias in favour of the southern summer cell. This cell is more extensive, and possesses a stream-function maximum more than twice the magnitude of the residual signature of the northern summer solstice, winter hemisphere cell. Further, the southern summer circulation is dominant regardless of the specific scheme used to treat atmospheric dust radiative heating and of whether the seasonal CO₂ cycle is active or not. The peak stream function in the annual average is roughly 25% of the magnitude of the instantaneous peak for the respective cells.

There are two main external factors which could influence the annual average tropical circulation: the eccentricity and the planetary topography. In particular, Mars possesses a significant zonal-average, north-to-south upward slope (this corresponds to an offset between the planetary centre-of-mass and centre-of-figure^{17,18}) and longitudinal asymmetry in topography¹⁸. Whether eccentricity or topography determines the asymmetry is critical. The former factor will change with the orbital elements, introducing climate variability that will also vary with the orbit, consistent with standard theory²⁻⁶. The latter would introduce a climatic asymmetry that is endogenic, emerging from the shape of the planet and not changing with the spin/orbit configuration. To isolate the eccentricity we reran the model with perihelion moved by 180° from its current

occurrence to $L_s = 70^\circ$ (see caption to Fig. 1 for definition of L_s). Figure 2b shows that the annual-average circulation in this case is little changed from that modelled for the present orbital state. Specifically, the tropical circulation remains dominated by the southern summer solstice, winter hemisphere cell, and cell shapes and stream-function magnitudes are not significantly different. The same result holds true for simulations with a circular orbit, and for simulations with no CO₂ or dust cycles. Eccentricity and the timing of perihelion are clearly not the dominant factors in the hemispheric biasing of the tropical circulation.

Speculation that topography might have an effect on the asymmetric strengths of the solstitial Hadley cells has generally focused on hemispheric differences in longitudinal structure^{12,19}. To isolate the influence of the zonal-mean topography, we undertook GCM simulations with the meridionally varying, zonal-mean component of elevation removed from the topography so that only longitudinal variations in topography of potential importance in the generation of waves and eddies remain. For the current timing of perihelion, Fig. 2c shows that without the dichotomy, the annual-average stream function is essentially symmetric about the equator. (Note that there remains some small hemispheric asymmetry in Fig. 2c that disappears if a smooth, thermally uniform surface is used. This slight deviation from hemispheric symmetry is due to the effect of the longitudinal topographic and surface thermal structure on the circulation.) This result also holds for non-dichotomy simulations run with perihelion at $L_s = 70^\circ$. We note that the dichotomy-free, northern hemisphere cell is significantly weaker in the annual average (by a factor of two in peak stream function) than in the case with real martian topography. The peak stream function for the

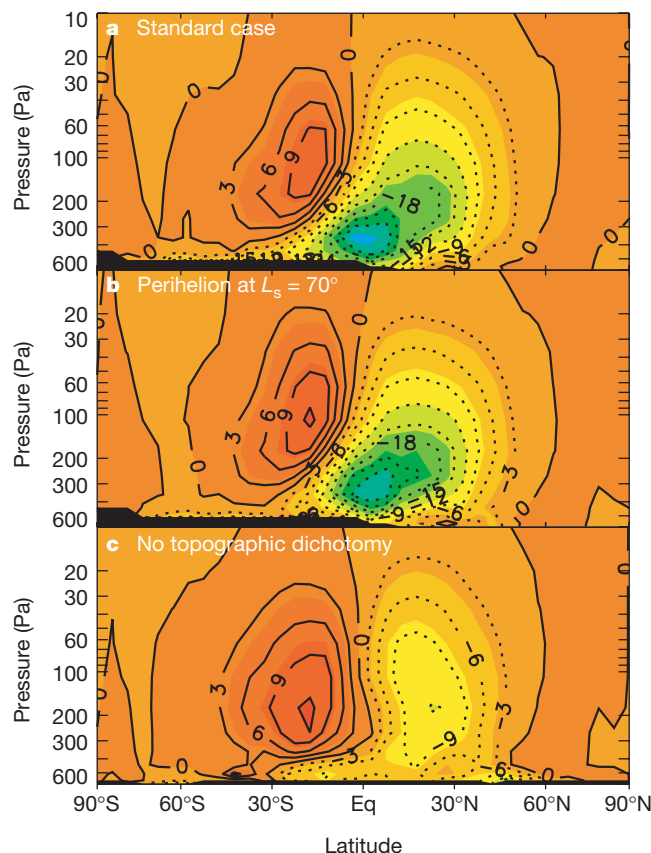


Figure 2 Mars GCM output demonstrating the annual-average bias in the tropical circulation, and that this bias is related to topography rather than the eccentric orbit of the planet. **a, b, c**, Annual-average, zonal-mean meridional stream functions are shown for **a**, observed topography and current orbital properties, **b**, observed topography, but the passage of perihelion moved to $L_s = 70^\circ$, and **c**, dichotomy-free topography, but current orbital properties. The stream functions are displayed in units of 10^8 kg s^{-1} , with positive

values indicating anticlockwise flow in the plane of the page. As in Fig. 1, the stream functions are the conventional eulerian mean. In these panels, the residual signature of the southern summer circulation is captured by the cell extending deep into the northern hemisphere, while the northern summer cell extends into the southern hemisphere. In each case, the dominant solstitial Hadley cell is the one that extends from the summer hemisphere into the winter hemisphere.

southern hemisphere cell is relatively unaffected by the change in topography. Simulations run with zonal-mean topography (not shown) result in a hemispheric bias in cell strength, as with the standard simulation.

These results demonstrate that the annual-average circulation is fundamentally biased by the zonal-mean topographic dichotomy on Mars. Elucidation of the specific mechanism(s) by which this asymmetry comes about requires further study, but probably involves the effect of the higher, southern hemisphere acting as an elevated heat source, and the effect of the cross-equatorial slope as a dynamical impediment to northern summer Hadley return flow. The former effect results from the fact that surface temperatures decrease much more slowly with pressure than do air temperatures. Thus elevated ground can significantly influence meridional gradients of temperature on pressure surfaces well above the ground, leading to biases in forcing of the tropical circulation, in analogy to the effect of the Tibetan plateau on the Indian monsoon²⁰.

The Hadley circulation has a strong influence on the cycles of dust and water, as described above. However, the evolution of water and dust within the climate system probably depends on a number of factors. For transport of water, the mixing capacity of the atmosphere must be convolved with the seasonal exchange of vapour with the seasonal and residual water ice caps. Water can also condense in the atmosphere at a level depending on the atmospheric temperature structure. If this level is sufficiently low, condensation will impede cross-equatorial transport^{21,22}. For dust mixing, the physics of dust injection at the surface is of major importance, and will depend on surface wind stresses and the temperature contrast between the surface and atmosphere across the boundary layer. Because dust is a radiatively active species there exists an important feedback between dust distribution, the distribution of radiative heating, and the circulation, with the circulation largely determining the dust distribution. Further, it is possible that the dust distribution is also influenced by the water cycle, through the use of dust particles as water cloud condensation nuclei.

Despite the complexity of the martian climate system, we can state that the asymmetry imposed on the climate by the topographic biasing of the Hadley circulation will have a uniquely important influence. Unlike the myriad other mechanisms operating within the system, which are determined by spatial patterns of solar heating, the topographic biasing of the Hadley circulation will impose a handedness upon the climate system that will remain as a residual when averaged over integer combinations of precessional cycles. While topography may also influence the CO₂ cycle on Mars²⁶, affecting the water cycle by varying the fraction of the year that water ice caps are exposed, the biasing of the circulation is far more direct and less susceptible to compensation by other processes (topographic biasing of the current CO₂ cycle would predict a permanent cap at the north—in reality, it is at the south). As such, this bias in the climate system is of great importance and requires further quantitative analysis.

We now consider how, and on what timescales, the topographic biasing will influence climate. The primary impact of the bias on the water cycle results from the net annual-average transport from the southern to northern hemispheres. Simulations using current orbital parameters, but with exposed water-ice caps prescribed at both poles, show that a southern residual water-ice cap is dynamically unstable with respect to the northern cap, even when cap temperatures are constrained to give no thermodynamic advantage to the north¹¹. It is possible that at high obliquity and with perihelion in northern summer, the northern cap will become unstable with respect to a southern residual water cap. However, it is equally clear that the net south-to-north transport generated by the dichotomy dictates that there will be more orbital configurations in which the northern cap, rather than the southern cap, will be stable.

Indeed, the existence of a southern residual water-ice cap may

require special combinations of eccentricity, obliquity and argument of perihelion. Asynchronous variation of these parameters (periods of 2 Myr, 0.12 Myr and 0.17 Myr, respectively²³) may imply that a southern water-ice cap occurs with a period of $>10^6$ years rather than the $\sim 10^5$ years predicted if one considers only cyclic variation in solar heating driven by precession. Polar layered deposit formation is poorly understood⁶, but it is widely held that layering develops at times of enhanced polar deposition of dust and water ice^{3–6}. If a residual water-ice cap is an important agent of layered deposit formation, then the tendency for water ice to be preferentially stable at the northern pole, and hence for a water ice cap to be much more likely to occur, suggests that the northern layered deposits should be more frequently active than those at the south, potentially by an order of magnitude or so. Clearly, the relationship between the topographic biasing of the water cycle and the disparate age of the layered deposits (with the south known to be much older than the north)²⁴ should be investigated in more detail.

Dark layers in the deposits are commonly ascribed to deposition during dustier climatic periods^{5,6}. The instantaneous strength of the Hadley circulation will play a large role in the bulk dustiness of the atmosphere by vertically and meridionally spreading dust. Models also suggest strong increases in surface winds as the Hadley cell intensifies^{7–9,12,13,25}, and these winds may contribute to lofting of dust, increasing the surface injection rate.

Although there are many configurations in which the northern summer solstitial Hadley cell may be dominant, the topographic bias dictates that the southern summer cell is more likely to be dominant, averaged over the cycles of precession, obliquity and eccentricity. As such, not only is it more likely that water ice will be stable at the northern pole and consequently that polar layered terrain formation is more common there, but the likelihood of southern summer being dustier than northern summer suggests that a ready supply of dust will be available to form layers. □

Received 9 October 2001; accepted 28 January 2002.

1. Zurek, R. W. *et al.* in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 835–933 (Univ. Arizona Press, Tucson, 1992).
2. Ward, W. R. Climatic variations on Mars. I. Astronomical theory of insolation. *J. Geophys. Res.* **97**, 3375–3386 (1974).
3. Murray, B. C., Ward, W. R. & Yeung, S. C. Periodic insolation variations on Mars. *Science* **180**, 638–640 (1973).
4. Kieffer, H. H. & Zent, A. P. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 1180–1220 (Univ. Arizona Press, Tucson, 1992).
5. Toon, O. B., Pollack, J. B., Ward, W., Burns, J. A. & Bilski, K. The astronomical theory of climate change on Mars. *Icarus* **44**, 552–607 (1980).
6. Thomas, P., Squyres, S. W., Herkenhoff, K., Howard, A. & Murray, B. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 767–798 (Univ. Arizona Press, Tucson, 1992).
7. Haberle, R. M., Leovy, C. B. & Pollack, J. B. Some effects of global dust storms on the atmospheric circulation of Mars. *Icarus* **50**, 322–367 (1982).
8. Wilson, R. J. & Hamilton, K. Comprehensive model simulation of thermal tide in the Martian atmosphere. *J. Atmos. Sci.* **53**, 1290–1326 (1996).
9. Murphy, J. R. *et al.* Three-dimensional numerical simulation of Martian global dust storms. *J. Geophys. Res.* **100**, 26357–26376 (1995).
10. Houben, H., Haberle, R. M., Young, R. E. & Zent, A. P. Modeling the Martian seasonal water cycle. *J. Geophys. Res.* **102**, 9069–9083 (1997).
11. Richardson, M. I. & Wilson, R. J. Investigation of the nature and stability of the Martian seasonal water cycle with a general circulation model. *J. Geophys. Res.* (in the press).
12. Haberle, R. M. *et al.* Mars atmospheric dynamics as simulated by the NASA Ames general-circulation model. I. The zonal-mean circulation. *J. Geophys. Res.* **98**, 3093–3123 (1993).
13. Forget, F. *et al.* Improved general circulation models of the Martian atmosphere from the surface to above 80 km. *J. Geophys. Res.* **104**, 24155–24175 (1999).
14. Oort, A. H. & Rasmussen, E. M. On the annual variation of the monthly mean meridional circulation. *Mon. Weath. Rev.* **98**, 423–442 (1970).
15. Peixoto, J. P. & Oort, A. H. *Physics of Climate* (American Institute of Physics, New York, 1992).
16. Lindzen, R. S. & Hou, A. Y. Hadley circulation for zonally averaged heating centered off the equator. *J. Atmos. Sci.* **45**, 2416–2427 (1988).
17. Smith, D. E. & Zuber, M. T. The shape of Mars and the topographic signature of the hemispheric dichotomy. *Science* **271**, 184–188 (1996).
18. Smith, D. E. *et al.* The global topography of Mars and implications for surface evolution. *Science* **284**, 1495–1503 (1999).
19. Haberle, R. M. *et al.* Mars general circulation model simulations with MOLA topography. *Icarus* (submitted).
20. Molnar, P. & Emanuel, K. A. Temperature profiles in radiative-convective equilibrium above surfaces at different heights. *J. Geophys. Res.* **104**, 24265–24271 (1999).
21. Clancy, R. T. *et al.* Water vapor saturation at low altitudes around Mars aphelion: A key to Mars climate? *Icarus* **122**, 36–62 (1996).

22. Richardson, M. I., Wilson, R. J. & Rodin, A. V. Water ice clouds in the Martian atmosphere: General circulation model experiments with a simple cloud scheme. *J. Geophys. Res.* (submitted).
23. Ward, W. R. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 298–320 (Univ. Arizona Press, Tucson, 1992).
24. Herkenhoff, K. E. & Plaut, J. J. Surface ages and resurfacing rates of the polar layered deposits on Mars. *Icarus* **144**, 243–253 (2000).
25. Fenton, L. K. & Richardson, M. I. Martian surface winds: Insensitivity to orbital changes and implications for aeolian processes. *J. Geophys. Res.* **106**, 32885–32902 (2001).
26. Jakosky, B. M., Henderson, B. G. & Mellon, M. T. The Mars water cycle at other epochs—Recent history of the polar caps and layered terrain. *Icarus* **102**, 286–297 (1993).

Acknowledgements

Discussions were provided by K. Emanuel, I. Held, A. Ingersoll, T. Schneider, and Y. Yung. We thank P. Gierasch for comments on the manuscript.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.I.R. (e-mail: mir@gps.caltech.edu).

Ferromagnetism in one-dimensional monatomic metal chains

P. Gambardella^{*,†}, A. Dallmeyer[†], K. Maiti[‡], M. C. Malagoli[‡], W. Eberhardt^{†,§}, K. Kern^{*,§} & C. Carbone^{†||}

^{*} Institut de Physique des Nanostructures, EPF-Lausanne, CH-1015 Lausanne, Switzerland

[†] Institut für Festkörperforschung, Forschungszentrum Jülich, D-52425 Jülich, Germany

[‡] Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung m.b.H., Albert-Einstein-Straße 15, 12489 Berlin, Germany

[§] Max-Planck-Institut für Festkörperforschung, Heisenbergstr. 1, D-70569 Stuttgart, Germany

^{||} Istituto di Struttura della Materia, Consiglio Nazionale delle Ricerche, Area Science Park, I-34012 Trieste, Italy

Two-dimensional systems, such as ultrathin epitaxial films and superlattices, display magnetic properties distinct from bulk materials¹. A challenging aim of current research in magnetism is to explore structures of still lower dimensionality^{2–6}. As the dimensionality of a physical system is reduced, magnetic ordering tends to decrease as fluctuations become relatively more important⁷. Spin lattice models predict that an infinite one-dimensional linear chain with short-range magnetic interactions spontaneously breaks up into segments with different orientation of the magnetization, thereby prohibiting long-range ferromagnetic order at a finite temperature^{7–9}. These models, however, do not take into account kinetic barriers to reaching equilibrium or interactions with the substrates that support the one-dimensional nanostructures. Here we demonstrate the existence of both short- and long-range ferromagnetic order for one-dimensional monatomic chains of Co constructed on a Pt substrate. We find evidence that the monatomic chains consist of thermally fluctuating segments of ferromagnetically coupled atoms which, below a threshold temperature, evolve into a ferromagnetic long-range-ordered state owing to the presence of anisotropy barriers. The Co chains are characterized by large localized orbital moments and correspondingly large magnetic anisotropy energies compared to two-dimensional films and bulk Co.

Since the work of Ising⁸, magnetism in one-dimensional (1D) systems has been the subject of continuous theoretical^{4,5,7–10} and experimental^{2,6,11} research. Progress in atomic engineering makes it possible today to build 1D arrays of transition-metal chains by self-

assembly epitaxial techniques on suitable substrates^{12,13}. Pioneering experiments^{14–16} in this direction investigated the magnetic behaviour of Fe stripes 1–10 nm wide obtained by step-flow growth on W(110) and Cu(111) surfaces. Ideally, one would like to construct monatomic chains in very large numbers while maintaining a fine control on the dimensions, uniformity and spatial distribution of the individual chains. We have shown¹⁷ that it is possible to produce high-density ($5 \times 10^6 \text{ cm}^{-1}$) arrays of parallel monatomic chains with unprecedented uniformity and even spacing by growing Co on a high-quality Pt vicinal surface in a narrow (250–300 K) temperature range. The scanning tunnelling microscopy (STM) images in Fig. 1 illustrate the regular step structure of the vicinal Pt(997) surface (Fig. 1a) and the fabrication of monatomic Co wires by decoration of the steps (Fig. 1b)¹⁷.

The magnetism of the Co wires has been investigated by X-ray magnetic circular dichroism¹⁸ (XMCD) at beamline ID-12B of the European Synchrotron Radiation Facility in Grenoble. Representative results for the circular dichroism in X-ray absorption spectroscopy (XAS) at the Co $L_{2,3}$ absorption edges for the monatomic wires are shown in Fig. 2a and compared with one monolayer (Fig. 2b) and bulk Co spectra (Fig. 2c). The difference between the two spectra for left- and right-handed polarization, reported in the lower panel, shows that the wires are characterized by a strong dichroism that results from the alignment of the Co magnetic moments in an external field of 7 T at 10 K. The amplitude of the dichroic signal is a measure of the magnetization of the Co wire array and contains information on the local character of the atomic moments.

The reduced atomic coordination of the monatomic chains compared to bulk Co and two-dimensional (2D) films has remarkable consequences for the relative size of the local orbital (m_L) and spin (m_S) magnetic moments. Both m_L and m_S are expected to increase as the atomic coordination is reduced in passing from the bulk to the monatomic chains. Local spin density calculations for Co/Pt(997) show that the narrowing of the Co 3d band and the

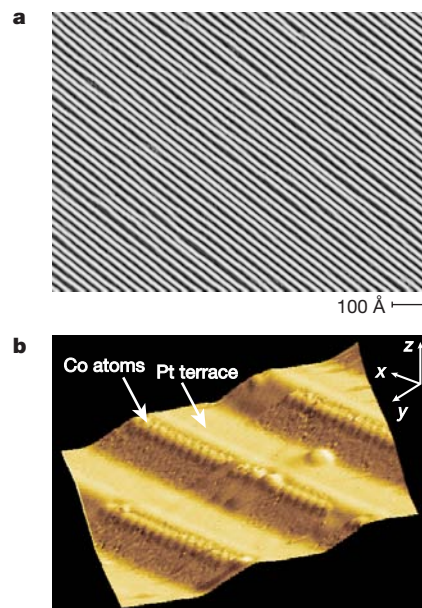


Figure 1 STM topographs of the Pt(997) surface. **a**, Periodic step structure (each white line represents a single step). The surface has a 6.45° miscut angle relative to the (111) direction; repulsive step interactions result in a narrow terrace width distribution centred at 20.2 Å with 2.9 Å standard deviation. **b**, Co monatomic chains decorating the Pt step edges (the vertical dimension is enhanced for better contrast). The monatomic chains are obtained by evaporating 0.13 monolayers of Co onto the substrate held at $T = 260 \text{ K}$ and previously cleaned by ion sputtering and annealing cycles in ultrahigh vacuum (UHV). The chains are linearly aligned and have a spacing equal to the terrace width.

corresponding increase in the density of states at the Fermi level produce the increase in m_s from the $1.57 \mu_B$ per atom bulk value to 2.03 and $2.08 \mu_B$ per atom for a monolayer and a 1D chain, respectively (G. Bihlmayer, X. Nie and S. Blügel, unpublished results); similar values can be found in refs 4 and 19). A larger relative increase is expected for m_L , which is generally more sensitive to changes in the atomic coordination because of its dependence on the crystal field. Using the XMCD sum rule²⁰, m_L can be determined from: $\int_{L_3+L_2} (\mu_+ - \mu_-) dE = (C/2\mu_B) m_L$, where μ_+ and μ_- are the absorption spectra for parallel and antiparallel direction of light polarization and magnetization, E is the photon energy and C is an experimental constant^{18,21} that is determined from the known bulk value $m_L = 0.14 \mu_B$ per atom²² (Fig. 2c). For the monatomic wires we find $m_L = 0.68 \pm 0.05 \mu_B$ per atom, an enhancement of about a factor of five compared to bulk Co. Dimensionality effects turn out to be decisive as we find that for biatomic wires and one monolayer Co m_L drops to $0.37 \mu_B$ per atom and $0.31 \pm 0.04 \mu_B$ per atom, respectively. The m_L value determined for the monatomic chains represents the highest orbital moment found in a 3d itinerant electron system, and is considerably larger than that of typical 2D Co structures²³ and nanoscale Co clusters²⁴.

In Fig. 3a we report the magnetic response of a set of monatomic wires at $T = 45$ K. The zero remanent magnetization (M) at $B = 0$ reveals the absence of long-range ferromagnetic order. However, the shape of the magnetization curve indicates the presence of short-range order, and therefore of significant interatomic exchange coupling in the chains. For non-interacting paramagnetic moments the magnetization expected in the present experimental conditions would be significantly smaller, as indicated by the dotted line in Fig. 3a. The observed behaviour is that of a 1D superparamagnetic system, that is, a system composed by segments, or spin blocks, each containing N exchange-coupled Co atoms, whose magnetization

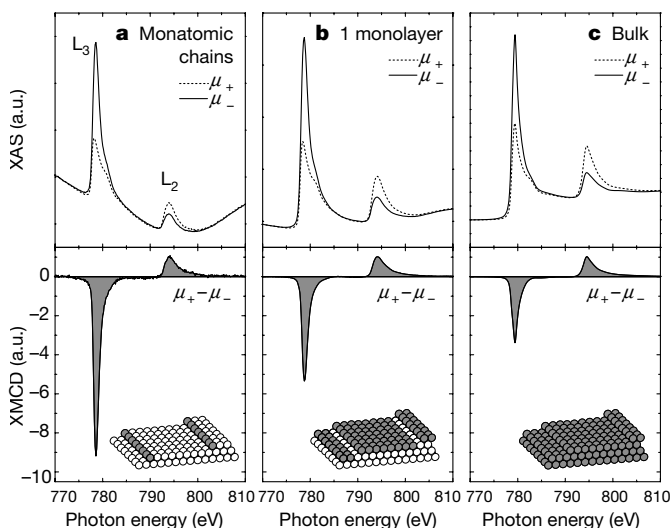


Figure 2 Co X-ray absorption spectra for parallel (μ_+) and antiparallel (μ_-) direction of light polarization and field-induced magnetization. The dichroism signal ($\mu_+ - \mu_-$) is obtained by subtraction of the absorption spectra in each panel and normalization to the L_2 peak. **a**, Monatomic chains; **b**, one monolayer; **c**, thick Co film on Pt(997). The sample was mounted onto a UHV variable-temperature insert that could be rotated with the respect to the direction of the external magnetic field applied parallel to the incident photon beam. Spectra were recorded in the electron-yield mode at $T = 10$ K and $B = 7$ T. Because of the low Co coverage, the edge structures of the monatomic wires are superimposed to a strong background originating from the oscillations following the Pt $N_{2,3}$ thresholds. As the structures from the non-magnetic substrate do not exhibit a dichroic effect, they do not contribute to the dichroic signal. Changes in the L_3 XMCD intensity indicate that the orbital moment is substantially increased in going from bulk Co to a 2D Co monolayer and, finally, to the 1D chains.

orientation is not stable owing to thermal fluctuations. A noticeable dependence of the magnetization on the direction of the applied field can be observed in Fig. 3. The strongest magnetic response is found in the direction pointing towards the step edges, perpendicular to the chain axis—at $+43^\circ$ from the (111) normal—as shown in the inset in Fig. 3a. Clearly, the shape of the superparamagnetic curves depends on the magnetic anisotropy energy of each spin block E_a , as well as on N times the magnetic moment per Co atom m , as in a paramagnetic system. According to classical Boltzmann statistics, the magnetization of an assembly of aligned particles with magnetic moment Nm and anisotropy E_a , placed in an external field B at an angle θ_0 with respect to the easy axis of magnetization is given by:

$$M = M_{\text{sat}} \frac{\int_0^{2\pi} d\varphi \int_0^\pi d\vartheta \sin \vartheta \cos \vartheta e^{(NmB \cos \vartheta + E_a(\sin \theta_0 \sin \vartheta \cos \varphi + \cos \theta_0 \cos \vartheta))/kT}}{\int_0^{2\pi} d\varphi \int_0^\pi d\vartheta \sin \vartheta e^{(NmB \cos \vartheta + E_a(\sin \theta_0 \sin \vartheta \cos \varphi + \cos \theta_0 \cos \vartheta))/kT}} \quad (1)$$

where ϑ and φ represent the spherical coordinates of Nm . By fitting the magnetization in Fig. 3a with equation (1), we obtain $Nm = 59 \pm 2 \mu_B$ and $E_a = 31 \pm 1$ meV. To correctly determine the number of coupled Co atoms per chain, N , we have to divide the calculated Nm value by $m = m_{\text{Co}} + m_{\text{Pt}}$, where $m_{\text{Co}} = m_L + m_s = 2.8 \mu_B$, and m_{Pt} represents the induced moment on the first- and second-nearest Pt neighbours. The latter amounts to about $1 \mu_B$ per Co atom, as derived on the basis of experimental²⁵ and theoretical¹⁹ results. Thus, we find that at 45 K about 15 Co atoms are coupled in each spin block, whereas the average length of a continuous Co chain (uninterrupted by kinks) is estimated to be about 80 atoms. The magnetic anisotropy energy (MAE) is 2.0 ± 0.2 meV per atom, a very large value compared to bulk h.c.p. Co ($40 \mu\text{eV}$ per atom²⁶) and compared to a Co monolayer on Pt(997) (0.14 ± 0.01 meV per atom). The MAE is directly related²¹ to the anisotropy of m_L measured in the easy and hard directions, which is unusually large ($0.12 \mu_B$) for the present system. The transition from 1D to a 2D system has profound effects on the magnetic anisotropy of the system: the MAE falls to 0.34 meV per atom for a bi-atomic chain and to 0.13 meV per atom for a Co monolayer film on Pt(997), a result that is related to the sharp decrease of the orbital magnetization reported above.

Here we cannot directly determine whether the measured size of the spin blocks is related to the occurrence of critical fluctuations in the 1D chains. We note, however, that the finite temperature and size of the chain system play a major role. A simple theoretical argument due to Landau⁷ shows that finite 1D Ising chains with a number of ferromagnetically-coupled spins $N > \exp(2J/kT)$, with J

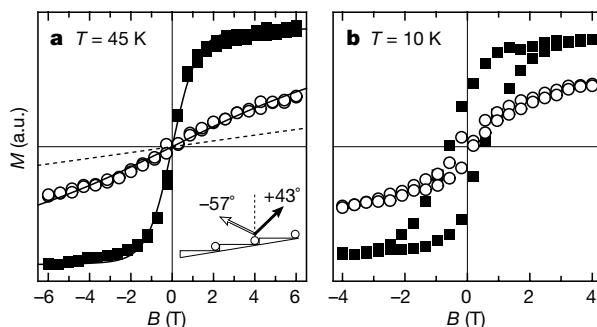


Figure 3 Magnetization of a monatomic wire array recorded at the L_3 Co edge. **a**, M as a function of the applied field at $T = 45$ K measured along the easy direction (filled squares) and at 80° away from the easy direction (open circles) in the plane perpendicular to the wire axis (see inset). The solid lines are fits to the data according to equation (1) with values $Nm = 59 \mu_B$, $E_a = 31$ meV, $T = 45$ K. The dashed line represents the magnetization expected for an isolated Co atom with $N = 1$, $m = 4 \mu_B$, $E_a = 2.1$ meV, $T = 45$ K. **b**, Hysteresis measured at $T = 10$ K below the blocking temperatures of the wires, in the same geometry as **a**.

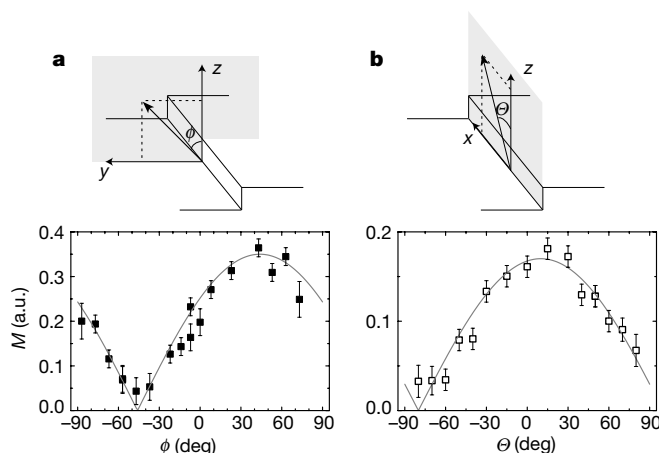


Figure 4 Angular dependence of the magnetization. **a**, Perpendicular; **b**, parallel to the wire axis at 10 K. For each angle the magnetization has been aligned in a 6 T magnetic field and subsequently measured near remanence at 0.25 T. A $|\cos(x - x_0)|$ dependence is found in both directions (solid lines), where $x = \phi, \theta$.

being the exchange energy between nearest-neighbour spins, break up spontaneously into smaller domains. Assuming the magnitude of $2J$ to be 15 meV (refs 6, 27), we get an upper limit of $N = 50$ atoms at $T = 45$ K. This argument, however, does not forbid the existence of ferromagnetism in finite 1D chains at finite temperature on the timescale of an experimental observation. As in bulk ferromagnetic systems, anisotropy energy barriers can effectively pin the magnetization along a fixed direction in space. Indeed, by lowering the sample temperature below 15 K, we observe a transition to a long-range ferromagnetically ordered state (Fig. 3b). Below the so-called blocking temperature ($T_b = 15 \pm 5$ K for the monatomic wires), the magnetization of each spin-block aligns in the easy axis direction, and the whole system becomes ferromagnetic. In a Néel–Brown model²⁸, the relaxation time of a single-domain magnetic particle is expressed by an Arrhenius law of the form $\tau = \tau_0 e^{-E_a/kT}$, where τ_0 is a prefactor of the order of 10^{-9} s. The determined anisotropy energy of $E_a = 31 \pm 1$ meV is thus consistent with the blocking temperature determined for the XMCD data, which yield a relaxation time $\tau \geq 10^2$ s at $T \leq 15$ K.

As in 2D films, the MAE and the easy axis of magnetization are determined by the wire atomic structure and by the substrate. Tight-binding MAE calculations of free-standing and Pd-deposited Co monatomic chains⁵ show that the easy direction rotates from parallel to the chain axis to out-of-plane in passing from the free-standing to the Pd-supported chains. Here the 1D geometry of the wires and the interaction with the substrate are manifested by a strong uniaxial anisotropic behaviour. Figure 4 shows the magnetization measured near remanence ($B = 0.25$ T) along different directions of the incident light with respect to the sample normal. M follows a cosine function both in the plane perpendicular to the wire axis and in the plane parallel to the wires, with a maximum in the perpendicular plane at $+43^\circ$. In our experiments we do not find evidence for interchain coupling effects¹¹ between the wires, either induced by the substrate or of dipolar origin. Ferromagnetic coupling between adjacent wires mediated by the polarization of the non-magnetic Pt host would result in an increase of the blocking temperature with respect to the value calculated by the Arrhenius law—such an increase is not observed. Interactions of dipolar origin, on the other hand, are far too weak (~ 0.06 meV between adjacent chains with perpendicular magnetization) compared to the MAE and also to thermal fluctuations ($kT \approx 4$ meV at 45 K and 0.8 meV at 10 K).

One-dimensional models are known to be useful in analysing many-body problems, but it is only recently that 1D systems made of real atoms have become the object of experiments. Here we have shown evidence that finite 1D 3d metal chains can sustain both

short- and long-range ferromagnetic order, and that parameters such as the size of the local magnetic moments and anisotropy are strongly dependent on the dimensionality of the Co chains. Future investigations could address the persistence and extent of short-range magnetic order in the chains as a function of their length and temperature. □

Received 17 August 2001; accepted 6 February 2002.

- Schneider, C. M. & Kirschner, J. in *Handbook of Surface Science* (eds Horn, K. & Scheffler, M.) 511–668 (Elsevier, Amsterdam, 2000).
- Himpfel, F. J., Ortega, J. E., Mankey, G. J. & Willis, R. F. Magnetic nanostructures. *Adv. Phys.* **47**, 511–597 (1998).
- Stamm, C. *et al.* Two-dimensional magnetic particles. *Science* **282**, 449–451 (1998).
- Weinert, M. & Freeman, A. J. Magnetism of linear chains. *J. Mag. Magn. Mater.* **38**, 23–33 (1983).
- Dorantes-Dávila, J. & Pastor, G. M. Magnetic anisotropy of one-dimensional nanostructures of transition metals. *Phys. Rev. Lett.* **81**, 208–211 (1998).
- Pratzer, M. *et al.* Atomic-scale magnetic domain walls in quasi-one-dimensional Fe nanostripes. *Phys. Rev. Lett.* **87**, 127201–1–127201–4 (2001).
- Landau, L. D. & Lifshitz, E. M. *Statistical Physics* Vol. 5, 482 (Pergamon, London, 1959).
- Ising, E. Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.* **31**, 253–258 (1925).
- Mermin, N. D. & Wagner, H. Absence of ferromagnetism in one- or two-dimensional isotropic Heisenberg models. *Phys. Rev. Lett.* **17**, 1133–1136 (1966).
- Lieb, E. H. & Mattis, D. C. *Mathematical Physics in One Dimension* (Academic, New York, 1966).
- De Jongh, L. J. & Miedema, A. R. Experiments on simple magnetic model systems. *Adv. Phys.* **23**, 1–260 (1974).
- Röder, H. *et al.* Building one- and two-dimensional nanostructures by diffusion-controlled aggregation at surfaces. *Nature* **366**, 141–143 (1993).
- Gambardella, P. *et al.* One-dimensional metal chains on Pt vicinal surfaces. *Phys. Rev. B* **61**, 2254–2262 (2000).
- Elmers, H. J. *et al.* Submonolayer magnetism of Fe(110) on W(110): finite width scaling of stripes and percolation between islands. *Phys. Rev. Lett.* **73**, 898–901 (1994).
- Hauschild, J., Elmers, H. J. & Gradmann, U. Dipolar superferromagnetism in monolayer nanostripes of Fe(110) nanostripes on vicinal W(110) surfaces. *Phys. Rev. B* **57**, R677–R680 (1998).
- Shen, J. *et al.* Magnetism in one dimension: Fe on Cu(111). *Phys. Rev. B* **56**, 2340–2343 (1997).
- Gambardella, P. *et al.* Co growth on Pt(997): from monatomic chains to monolayer completion. *Surf. Sci.* **449**, 93–103 (2000).
- Stöhr, J. & Nakajima, R. Magnetic properties of transition-metal multilayers studied with XMCD spectroscopy. *IBM J. Res. Develop.* **48**, 73–88 (1998).
- Wu, R., Li, C. & Freeman, A. J. Structural, electronic and magnetic properties of Co/Pd(111) and Co/Pt(111). *J. Mag. Magn. Mater.* **99**, 71–80 (1991).
- Thole, B. T., Cara, P., Sette, F. & van der Laan, G. X-ray circular dichroism as a probe of orbital magnetization. *Phys. Rev. Lett.* **68**, 1943–1946 (1992).
- Weller, D. *et al.* Microscopic origin of magnetic anisotropy in Au/Co/Au probed with x-ray magnetic circular dichroism. *Phys. Rev. Lett.* **75**, 3752–3755 (1995).
- Chen, C. T. *et al.* Experimental confirmation of the XMCD sum rules for iron and cobalt. *Phys. Rev. Lett.* **75**, 152–155 (1995).
- Tischer, M. *et al.* Enhancement of orbital magnetism at surfaces: Co on Cu(100). *Phys. Rev. Lett.* **75**, 1602–1605 (1995).
- Dürr, H. A. *et al.* Spin and orbital magnetization in self-assembled Co clusters on Au(111). *Phys. Rev. B* **59**, R701–R704 (1999).
- Crangle, J. & Scott, W. R. Dilute ferromagnetic alloys. *J. Appl. Phys.* **36**, 921–928 (1965).
- Bate, G. Magnetic recording materials since 1975. *J. Mag. Magn. Mater.* **100**, 413–424 (1991).
- Fróta-Pessoa, S., Muniz, R. B. & Kudrnovský, J. Exchange coupling in transition-metal ferromagnets. *Phys. Rev. B* **62**, 5293–5296 (2000).
- Wernsdorfer, W. *et al.* Experimental evidence of the Néel–Brown model of magnetization reversal. *Phys. Rev. Lett.* **78**, 1791–1794 (1997).

Acknowledgements

We thank P. Ohresser, S. S. Dhesi, K. Larsson and N. B. Brookes of beamline ID12B of the European Synchrotron Radiation Facility in Grenoble for their assistance during the experiment.

Correspondence and requests for materials should be addressed to P.G. (e-mail: pietro.gambardella@epfl.ch).

An ordered mesoporous organosilica hybrid material with a crystal-like wall structure

Shinji Inagaki*, Shiyu Guan[†], Tetsu Ohsuna[‡] & Osamu Terasaki[§]

* Toyota Central R&D Laboratories, Inc., Nagakute, Aichi, 480-1192, Japan

[‡] Institute for Materials Research, Tohoku University, Sendai 980-8577, Japan

[§] Department of Physics and CREST, JST, Graduate School of Science, Tohoku University, Sendai 980-8578, Japan

[†] Present address: Sanyo Chemical Industry, Ltd, Higashiyama, Kyoto, 605-0995, Japan

Surfactant-mediated synthesis strategies are widely used to fabricate ordered mesoporous solids^{1–6} in the form of metal oxides⁷, metals⁸, carbon⁹ and hybrid organosilicas^{10–14}. These materials have amorphous pore walls, which could limit their practical utility. In the case of mesoporous metal oxides, efforts to crystallize the framework structure by thermal^{15,16} and hydrothermal treatments¹⁷ have resulted in crystallization of only a fraction of the pore walls. Here we report the surfactant-mediated synthesis

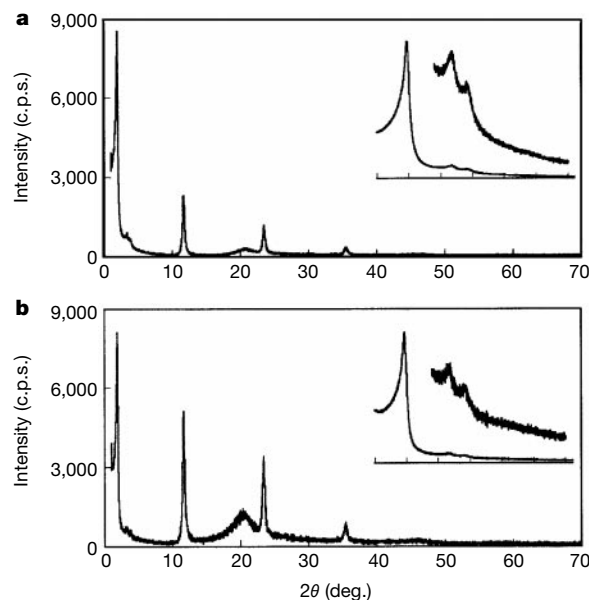


Figure 1 Powder X-ray diffraction patterns of mesoporous benzene-silicas. **a**, Material after removal of surfactants. **b**, As-made material containing surfactants. Patterns in the low-angle region ($1 < 2\theta < 7^\circ$) are shown magnified in the insets. These materials have both mesoscale ($d = 45.5$, 26.0 and 22.9 Å) and molecular-scale ($d = 7.6$, 3.8 and 2.5 Å) periodic structures.

of an ordered benzene-silica hybrid material; this material has an hexagonal array of mesopores with a lattice constant of 52.5 Å, and crystal-like pore walls that exhibit structural periodicity with a spacing of 7.6 Å along the channel direction. The periodic pore

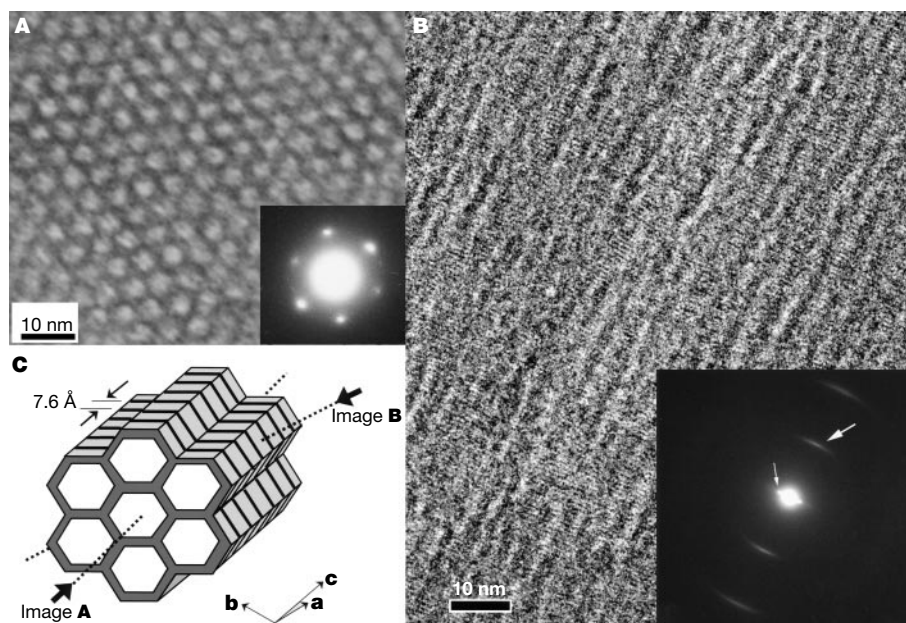


Figure 2 TEM images, electron diffraction patterns and the resulting structural model of mesoporous benzene-silica. The images and the patterns are arranged in the correct orientation relation—that is, the diffraction spots and corresponding lattice planes are normal to each other. **A**, Image and pattern taken with $[001]$ incidence, parallel to the channels. Uniform mesopores with a diameter of 38 Å are arranged in a hexagonal manner. **B**, Image and pattern taken with $[100]$ incidence, perpendicular to the channels. Many lattice fringes with a spacing of 7.6 Å are observed in the pore walls. The wavy contrast, which is perpendicular to the lattice fringes, with a spacing of 45.5 Å

($d = \sqrt{3}a/2$) are also discernible. Note that we cannot observe the contrast of 7.6 -Å and 45.5 -Å spacings at the best condition simultaneously for both, because the dependence of the contrast transfer function of the objective lens on focus condition is different for both. The electron diffraction pattern also shows diffused spots due to the 7.6 -Å periodicity (large arrow) in the perpendicular direction to the spots due to channel arrangement with $d = 45.5$ Å (small arrow). **C**, Schematic model of mesoporous benzene-silica derived from the results of the TEM images and electron diffraction patterns.

surface structure results from alternating hydrophilic and hydrophobic layers, composed of silica and benzene, respectively. We believe that this material is formed as a result of structure-directing interactions between the benzene–silica precursor molecules, and between the precursor molecules and the surfactants. We expect that other organosilicas and organo-metal oxides can be produced in a similar fashion, to yield a range of hierarchically ordered mesoporous solids with molecular-scale pore surface periodicity.

To produce the material, the benzene-bridged organosilane monomer, $(\text{C}_2\text{H}_5\text{O})_3\text{Si}-\text{C}_6\text{H}_4-\text{Si}(\text{OC}_2\text{H}_5)_3$ (1,4-bis(triethoxysilyl)-benzene, BTEB), was added to an aqueous solution of alkyltrimethylammonium surfactant (3.3 wt%) containing sodium hydroxide, and kept at 95 °C for about 20 hours. It is important to ensure that the solution is alkaline, and the mixture ratio optimized. The benzene–silica hybrid material was obtained by collecting the white precipitate (as-made materials), and removing surfactant by solvent extraction. The material is in the form of plate-like particles (0.5–30 μm in side length and 1 μm in thickness), composed of needle-like single crystals aligned perpendicular to the plate surface (see Supplementary Information).

The powder X-ray diffraction (XRD) pattern of the benzene–silica hybrid material (surfactant-free) showed three peaks in the small-angle scattering regime ($2\theta < 10^\circ$), with spacings d of 45.5, 26.0 and 22.9 Å, which can be indexed to a two-dimensional

hexagonal ($p6mm$) lattice with a lattice constant $a = 52.5$ Å (Fig. 1a inset). The as-made material, containing surfactant, showed a similar XRD pattern, with the somewhat larger lattice constant $a = 53.1$ Å (Fig. 1b inset). Both transmission electron microscope (TEM) images and corresponding electron diffraction patterns revealed a clear hexagonal arrangement of one-dimensional channels with uniform size (Fig. 2A and inset). Nitrogen adsorption isotherms also confirmed the existence of uniform mesopores, with the modified BJH (Barrett–Joyner–Halenda) pore diameter¹⁸ of 38 Å. The BET (Brunauer–Emmett–Teller) surface area and mesopore volume were $818\text{ m}^2\text{ g}^{-1}$ and $0.66\text{ cm}^3\text{ g}^{-1}$, respectively (see Supplementary Information).

The XRD patterns at medium scattering angles ($2\theta = 10\text{--}70^\circ$) display four sharp peaks at $d = 7.6, 3.8, 2.5$ and 1.9 Å (in addition to three peaks at small scattering angles) (Fig. 1a). These diffraction peaks can be explained by a periodic structure with a spacing of 7.6 Å. TEM images reveal many lattice fringes—stacked along the channel axes, with a uniform spacing of 7.6 Å—on the pore walls over the whole region (Fig. 2B). Electron diffraction spots corresponding to 7.6-Å periodicity and its higher order are observed perpendicular to the diffraction spot owing to the mesopore arrangement (with $a = 52.5$ Å), although we cannot observe higher-order reflections of 45.5-Å periodicity (Fig. 2B inset). These results clearly indicate that molecular-scale periodicity exists in the whole region of the pore walls, which consist of the

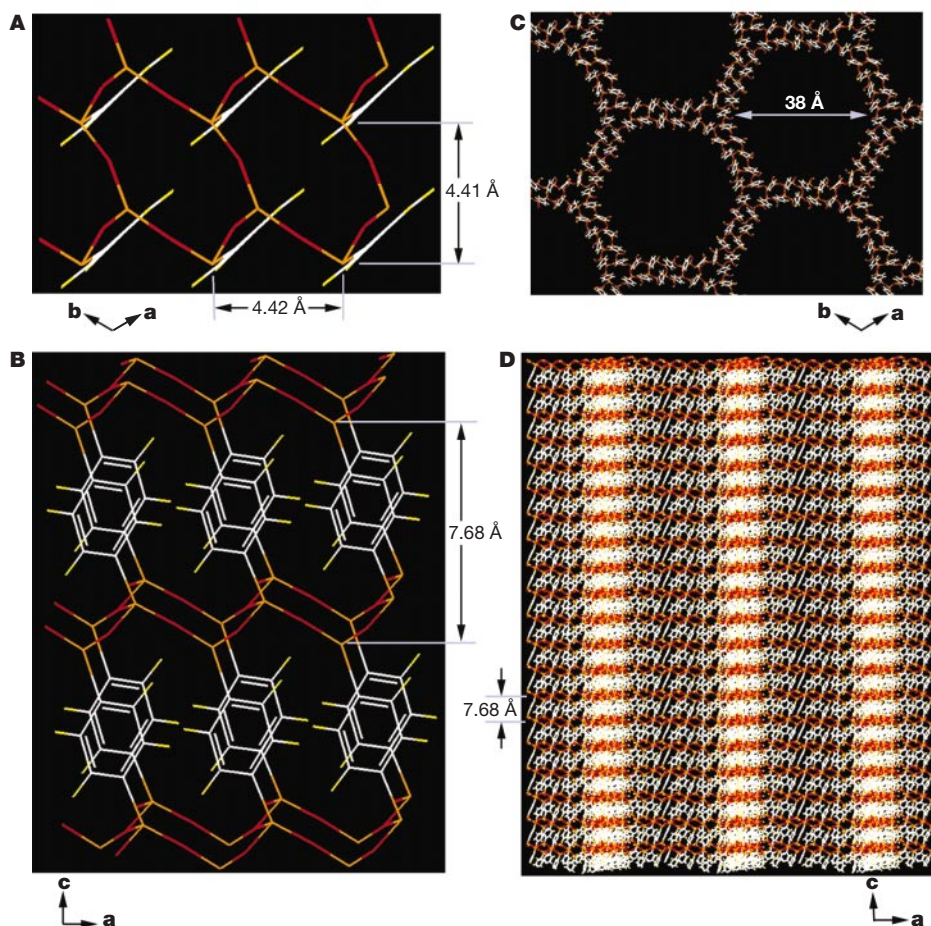


Figure 3 Structural models of mesoporous benzene–silica. **A, B**, Images of the layered arrangement of $\text{SiO}_{1.5}-\text{C}_6\text{H}_4-\text{SiO}_{1.5}$ units in the walls. The structure was optimized by minimizing the three-dimensional periodic lattice using the force field COMPASS. **C, D**, Images of the hexagonal lattice constructed with the layered pore-wall structure.

The structure was also minimized by using the force field COMPASS. Atoms are represented as a stick model. Silicon, orange; oxygen, red; carbon, white; hydrogen, yellow.

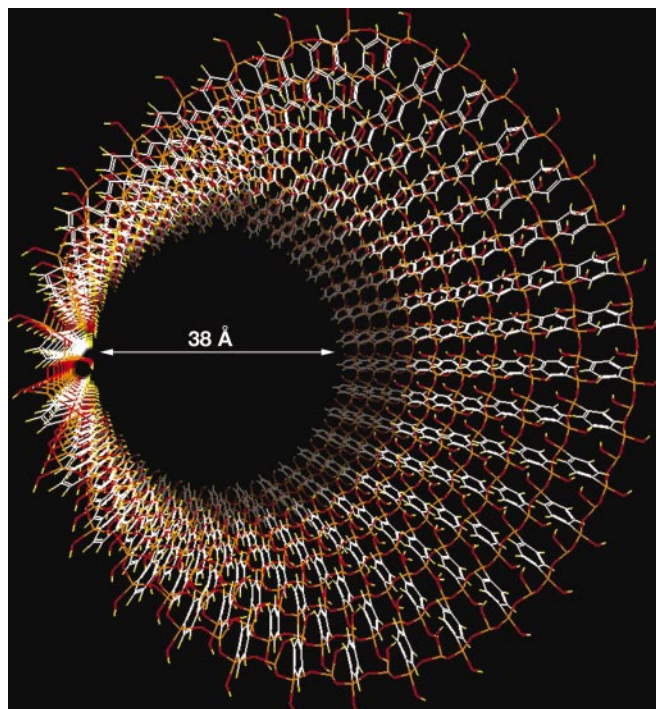


Figure 4 Model showing the pore surface of mesoporous benzene-silica. Benzene rings are aligned in a circle around the pore, fixed at both sides by silicate chains. The silicate is terminated by silanol (Si-OH) at the surface. Hydrophobic benzene layers and hydrophilic silicate layers array alternately at an interval of 7.6 Å along the channel direction. Silicon, orange; oxygen, red; carbon, white; hydrogen, yellow.

benzene-silica composite. As-made mesoporous benzene-silica also displays the 7.6-Å periodicity along the walls (Fig. 2C). The peak intensity of the 7.6-Å periodicity in the powder XRD pattern is almost the same for surfactant-free and as-made mesoporous benzene-silicas, indicating that the periodicity is completely retained after removal of surfactant (Fig. 1b). ^{29}Si and ^{13}C magic angle spinning nuclear magnetic resonance (MAS NMR) measurements clearly showed that the pore walls of the mesoporous benzene-silica are made of a covalently bonded network composed of $\text{O}_{1.5}\text{Si}-\text{C}_6\text{H}_4-\text{SiO}_{1.5}$ units, and that no carbon-silicon bond cleavage of the BTEB molecule occurred during the synthesis (Supplementary Information). The sharp signals due to silicon species of the ^{29}Si MAS NMR spectra are indicative of a unique and uniform environment around Si atoms in the walls, consistent with the crystal-like pore-wall structure found in XRD, electron diffraction and TEM observations.

Figure 3 shows a structural model of the mesoporous benzene-silica, which is constructed on the basis of the well-defined structure of crystalline 1,4-bis(trihydroxysilyl)benzene ((HO) $_3\text{Si}-\text{C}_6\text{H}_4-\text{Si}(\text{OH})_3$, BTHB; ref. 19). The BTHB molecule, a triol analogue of the BTEB molecule, forms a layered structure in which BTHB molecules are packed in a head-to-tail manner within sheets with an interlayer spacing of 10.1 Å. Hydrogen-bonding of silanols ($\text{Si}-\text{OH}\cdots\text{HO}-\text{Si}$) among BTHB molecules stabilizes the crystal structure. We substituted covalent bonds ($\text{Si}-\text{O}-\text{Si}$) for the hydrogen bonds in the model of mesoporous benzene-silica (Fig. 3A and B). Minimizing the framework energy of the covalently bonded periodic lattice using a molecular mechanics simulation resulted in a slight decrease in the interlayer spacing, to 7.68 Å, in good agreement with the 7.6-Å periodicity observed by XRD and electron diffraction patterns. We have constructed a hexagonal lattice model of the mesoporous structure on the basis of the periodic pore-wall structure (Fig. 3C and D). The side view of the pore walls in the model (Fig. 3D) resembles very closely the images of TEM (Fig. 2B),

supporting the validity of the structural model. The mesoporous benzene-silica is thus a hierarchically ordered porous material, displaying both meso- and molecular-scale periodicity. The material had extremely high thermal and hydrothermal stability: thermogravimetric analysis showed that benzene groups were kept in the walls up to 500 °C in air or nitrogen (Supplementary Information), and the meso- and molecular-scale periodicity was completely preserved even after boiling the material in water for 8 hours.

Self-assembly of the framework source of organosilane BTEB molecules formed the periodic structure in the walls of the mesoporous benzene-silica probably because hydrophobic and hydrophilic interactions directed the self-assembly of BTEB molecules into the framework structure¹⁹. Similar molecular-scale periodicity was also observed in mesoporous ethylene ($-\text{CH}=\text{CH}-$)-silica, though it had smaller molecular-scale periodicity (5.6 Å). The periodicity in the walls increased with increasing molecular length of organic groups, suggesting that the layered structure with the organic group as a pillar is essential for formation of the organosilica hybrid materials. In the case of smaller interactive organic groups of ethane ($-\text{CH}_2-\text{CH}_2-$) and methane ($-\text{CH}_3-$), the mesoporous materials showed no molecular-scale periodicity, even though they had highly ordered mesostructures. This suggests that strong interaction of organosilane molecules is important to form the molecular-scale ordered structures. However, molecular-scale periodicity is not seen in the non-periodic amorphous benzene-silica and other organic-inorganic hybrid materials produced by sol-gel methods not using surfactants^{20,21}. This suggests that the self-assembly of the surfactant and the resulting mesostructure promotes the formation of the molecular-scale periodic wall structure. These observations suggest that analogous interactions can be exploited in the synthesis of a variety of hierarchically ordered materials with not only hybrid organosilicas but also hybrid organo-metal oxides.

We find a unique surface structure in the model of mesoporous benzene-silica, in which hydrophilic silicate layers and hydrophobic benzene layers array alternately (Fig. 4). This structure, with periodically arranged hydrophobic-hydrophilic surfaces, could enable structural orientation of guest molecules or clusters enclosed in the pores, which in turn might improve the selectivity and activity in catalytic applications, and enhance the opto-electrical efficiency of included molecules and clusters. We have sulphonated the mesoporous benzene-silica while preserving both the meso- and molecular-scale ordered structures (see Supplementary Information). The sulphonated acid-functionalized mesoporous material might find use as a solid acid catalyst that can be readily recovered and reused²². The sulphonate groups were kept at pore walls up to 500 °C in air or nitrogen, as confirmed by mass spectroscopy. This unusually high thermal stability would allow the use of this solid acid catalyst not only in liquid-phase reactions, but also in gas-phase reactions at high temperature. Because the sulphonated acid groups serve as carriers of protons, the sulphonated mesoporous benzene-silica might also find use as an electrolyte for fuel cells. Moreover, fluorescence spectra indicate some interactions between benzene rings in the walls (see Supplementary Information). The closest intermolecular distance of benzene rings is 4.4 Å in the model (Fig. 3A); if closer staking of benzene rings in the walls could be achieved, $\pi-\pi$ conjugation of the rings would render the porous framework conducting. □

Methods

Preparation of mesoporous benzene-silica material

Mesoporous benzene-silica was synthesized from 1,4-bis(trihydroxysilyl)benzene (BTEB), Azmax Japan, in the presence of octadecyltrimethylammonium chloride (ODTMA) surfactant. ODTMA (16.665 g, 47.88 mmol) was dissolved in a mixture of ion-exchanged water (500 g) and 6 M sodium hydroxide (NaOH) aqueous solution (40 g, 200 mmol NaOH) at 50–60 °C. BTME (20 g, 49.67 mmol) was added to the ODTMA solution under vigorous stirring at room temperature. The mixture was treated ultrasonically for 20 min

to disperse the hydrophobic BTEB in the aqueous solution, and stirred for 20 h at room temperature. The solution was kept at 95 °C for 20 h under static conditions. The resulting white precipitate was recovered by filtration and drying to yield as-made mesoporous benzene–silica material (8.22 g). Surfactant was removed by stirring 1.0 g of as-synthesized material in 250 ml of ethanol with 9 g of 36% HCl aqueous solution at 70 °C for 8 h to yield mesoporous benzene–silica (0.69 g).

Sulphonation of mesoporous benzene–silica

The mesoporous benzene–silica was dried under evacuation at 10^{-2} – 10^{-3} torr for 4 h and added to 25% $\text{SO}_3/\text{H}_2\text{SO}_4$ solution. The dispersed solution was kept at 105–110 °C for 5 h. The dispersion was cooled at room temperature, and added to a large volume of water. The solid was filtered and washed with water repeatedly, then boiled in ion-exchanged water for 1 h, filtered and washed with boiling water repeatedly. Finally, the solid was stirred in 6 M hydrochloric acid aqueous solution for 24 h to protonate and form sulphonic acid groups. The acid content of the mesoporous material was determined from the titration curve. The sulphonated sample (50 mg) was immersed in 10 wt% sodium chloride aqueous solution for 24 h. The solution was titrated with 0.05 M NaOH to produce the titration curve. The acid amount was determined to be 0.40 mequiv. g^{-1} . The sulphonation conditions could sulphonate both the benzene rings and the silica.

Received 1 November 2001; accepted 22 January 2002.

- Yanagisawa, T., Shimizu, T., Kuroda, K. & Kato, C. The preparation of alkyltrimethylammonium-kanemite complexes and their conversion to microporous materials. *Bull. Chem. Soc. Jpn* **63**, 988–992 (1990).
- Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
- Beck, J. S. *et al.* A new family of mesoporous molecular sieves prepared with liquid crystal templates. *J. Am. Chem. Soc.* **114**, 10834–10843 (1992).
- Inagaki, S., Fukushima, Y. & Kuroda, K. Synthesis of highly ordered mesoporous materials from a layered polysilicate. *J. Chem. Soc. Chem. Commun.* 680–682 (1993).
- Firouzi, A. *et al.* Cooperative organization of inorganic-surfactant and biomimetic assemblies. *Science* **267**, 1138–1143 (1995).
- Bagshaw, S. A., Prouzet, E. & Pinnavaia, T. J. Templating of mesoporous molecular sieves by nonionic polyethylene oxide surfactants. *Science* **269**, 1242–1244 (1995).
- Ying, J. Y., Mehnert, C. P. & Wong, M. S. Synthesis and applications of supramolecular-templated mesoporous materials. *Angew. Chem. Int. Edn Engl.* **38**, 56–77 (1999).
- Attard, G. S. *et al.* Mesoporous platinum films from lyotropic liquid crystalline phases. *Science* **278**, 838–840 (1997).
- Joo, S. H. *et al.* Ordered nanoporous arrays of carbon supporting high dispersions of platinum nanoparticles. *Nature* **412**, 169–172 (2001).
- Inagaki, S., Guan, S., Fukushima, Y., Ohsuna, T. & Terasaki, O. Novel mesoporous materials with a uniform distribution of organic groups and inorganic oxide in their framework. *J. Am. Chem. Soc.* **121**, 9611–9614 (1999).
- Melda, B. J., Holland, B. T., Blanford, C. F. & Stein, A. Mesoporous sieves with unified hybrid inorganic/organic framework. *Chem. Mater.* **11**, 3302–3308 (1999).
- Asefa, T., MacLachlan, M. J., Coombs, N. & Ozin, G. A. Periodic mesoporous organosilicas with organic groups inside the channel walls. *Nature* **402**, 867–871 (1999).
- Guan, S., Inagaki, S., Ohsuna, T. & Terasaki, O. Cubic hybrid organic-inorganic mesoporous crystal with a decaoctahedral shape. *J. Am. Chem. Soc.* **122**, 5660–5661 (2000).
- Stein, A., Melde, B. J. & Schroden, R. C. Hybrid inorganic-organic mesoporous silicates—Nanoscale reactors coming of age. *Adv. Mater.* **12**, 1403–1419 (2000).
- Yang, P., Zhao, D., Margolese, D. I., Chmelka, B. F. & Stucky, G. D. Generalized syntheses of large-pore mesoporous metal oxides with semicrystalline frameworks. *Nature* **396**, 152–155 (1998).
- Lee, B., Lu, D., Kondo, J. N. & Domen, K. Single crystal particles of a mesoporous mixed transition metal oxide with a wormhole structure. *Chem. Commun.* 2118–2119 (2001).
- Liu, Y., Zhang, W. & Pinnavaia, T. J. Steam-stable MSU-S aluminosilicate mesostructures assembled from zeolite ZSM-5 and zeolite beta seeds. *Angew. Chem. Int. Edn Engl.* **40**, 1255–1258 (2001).
- Kruk, M. & Jaroniec, M. Gas adsorption characterization of ordered organic-inorganic nanocomposite materials. *Chem. Mater.* **13**, 3169–3183 (2001).
- Cerveau, G., Corriu, R. J., Dabien, B. & Bideau, J. L. Synthesis of stable organo(bis-silanetriols): X-ray powder structure of 1,4-bis(trihydroxysilyl)benzene. *Angew. Chem. Int. Edn Engl.* **39**, 4533–4537 (2000).
- Loy, D. A. & Shea, K. J. Bridged polysilsesquioxanes—Highly porous hybrid organic-inorganic materials. *Chem. Rev.* **95**, 1431–1442 (1995).
- Corriu, R. J. P. Ceramics and nanostructures from molecular precursors. *Angew. Chem. Int. Edn Engl.* **39**, 1376–1398 (2000).
- Harmer, M. A., Farneth, W. E. & Sun, Q. Towards the sulfuric acid of solids. *Adv. Mater.* **10**, 1255–1257 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank S. Yamamoto for help in computer simulation of structural models, and S. Hyde for critical reading of the manuscript. O.T. was supported by CREST, Japan Science and Technology Corporation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.I. (e-mail: inagaki@mosk.tytlabs.co.jp).

Origin and fate of Lake Vostok water frozen to the base of the East Antarctic ice sheet

Robin E. Bell*, Michael Studinger*, Anahita A. Tikku*, Garry K.C. Clarke†, Michael M. Gutner‡ & Chuck Meertens§

* Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

† Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

‡ Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06511, USA

§ UNAVCO/UCAR Facility, Boulder, Colorado 80307, USA

The subglacial Lake Vostok may be a unique reservoir of genetic material and it may contain organisms with distinct adaptations^{1–3}, but it has yet to be explored directly. The lake and the overlying ice sheet are closely linked, as the ice-sheet thickness drives the lake circulation, while melting and freezing at the ice-sheet base will control the flux of water, biota and sediment through the lake^{4–7}. Here we present a reconstruction of the ice flow trajectories for the Vostok core site, using ice-penetrating

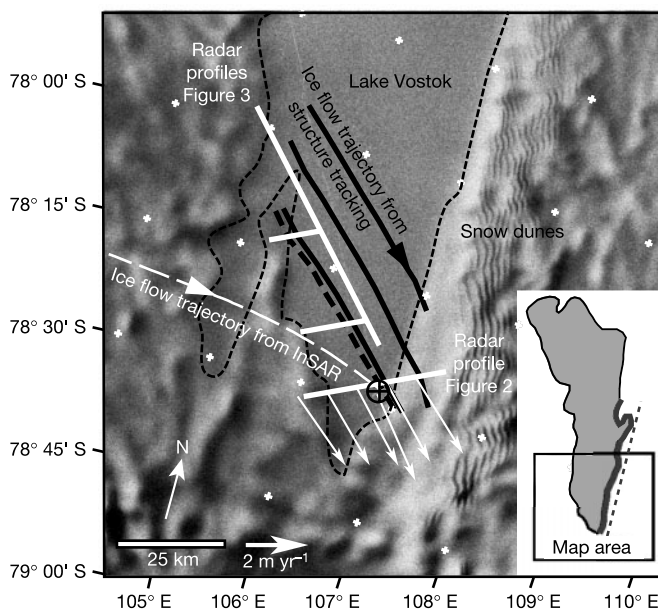


Figure 1 RADARSAT synthetic aperture radar image¹⁵ of the ice sheet surface over the southern part of Lake Vostok. Inset, map showing the outline of the entire lake, indicating the location of the enlarged ice-surface image. In the inset, the heavy line is the region where accreted ice is observed leaving the lake, and the dashed line is the downslope line where the ice flux is estimated. In the RADARSAT image, the grounding line is marked by the fine dashed line between the smooth, featureless ice surface over the lake and the rough ice surface over the grounded ice sheet. The shallow, narrow embayment and its possible connection to the main lake are seen along the western shoreline. The ice flow trajectory shown white is derived from interferometry^{4,12}. The four ice flow trajectories derived from internal layer structures are shown as black lines across the lake. The Vostok flow trajectory is illustrated as a dashed black line. The Vostok ice-core site is marked by the circle. The location of the radar profile shown in Fig. 2 is indicated by the long white line that crosses the Vostok core site. The radar profiles shown in the 'fence diagram' (Fig. 3) are shown as the intersecting white lines to the northwest. The GPS velocity vectors are shown as arrows.

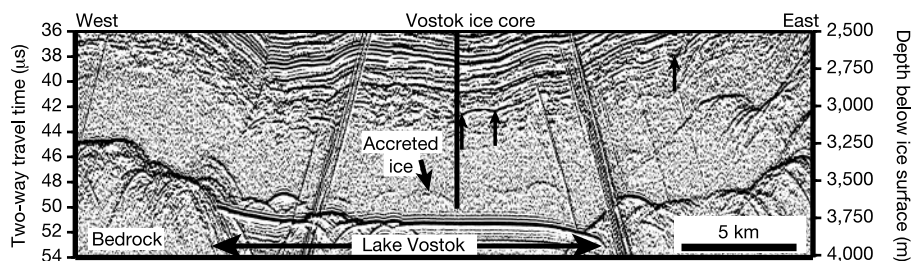


Figure 2 Airborne ice-penetrating radar data through the Vostok core site transverse to the ice flow. The internal layers are the rugged, roughly horizontal features in the upper half of the image. The lake is the strong flat reflector between 50 and 51 μs two-way travel time (3,745 m below the ice surface at the Vostok core site). The ice sheet sags as it enters the lake on the western grounding line, and compresses as it flows out to the east. The bounding topography is the rough feature on the western and eastern side of the

image. The accreted ice layer is noted at 49 μs —3,570 m below the surface. The accreted ice layer is observed to the east of the lake between the base of the internal reflectors and the subglacial topography. Three of the features in the internal layers traced to generate the flowlines are indicated with arrows. Vostok station appears as a major surface diffractor, roughly centred over the core site.

radar data and Global Positioning System (GPS) measurements of surface ice velocity. We find that the ice sheet has a significant along-lake flow component, persistent since the Last Glacial Maximum. The rates at which ice is frozen (accreted) to the base of the ice sheet are greatest at the shorelines, and the accreted ice layer is subsequently transported out of the lake. Using these new flow field and velocity measurements, we estimate the time for ice to traverse Lake Vostok to be 16,000–20,000 years. We infer that most Vostok ice analysed to date was accreted to the ice sheet close to the western shoreline, and is therefore not representative of open lake conditions. From the amount of accreted lake water we estimate to be exported along the southern shoreline, the lake water residence time is about 13,300 years.

In the 2000–01 field season, an integrated aerogeophysical survey that included ice-penetrating radar soundings (J. W. Holt *et al.*, manuscript in preparation) was conducted over Lake Vostok (Fig. 1). The ice-penetrating radar profile that passes closest to the Vostok ice core, located 1.5 km north of the core site, shows an ice thickness of $3,745 \pm 16$ m using a radar velocity in ice⁸ of $168.5 \text{ m } \mu\text{s}^{-1}$; this thickness is similar to earlier estimates⁹ ($3,750 \pm 30$ m). These radar data are characterized by isochronous internal layers, which result

from the changing conductivity associated with periods of increased precipitation of volcanic dust^{10,11}. Radar data at the core site reveal continuous internal layers down to 2,900 m, below which there are no coherent parallel structures (Fig. 2). At 3,570 m we detect a distinctive reflector across the width of the lake; this reflector is absent to the west of the lake, but present in the grounded ice downflow from the lake. Features in the reflector can be traced in adjacent radar lines, indicating that the reflector is the result of physical changes in the base of the ice sheet and not the result of isolated englacial debris or hyperbolic reflections from subglacial topography. Samples of chemically and physically distinct ice recovered below 3,540 m in the Vostok core represent the upper fraction of a ~200-m-thick basal ice layer derived from accretion of lake water onto the base of the ice sheet⁵. We interpret this reflector as corresponding to the level in the Vostok ice core where solid mineral inclusions are highly concentrated. As the accretion reflector is present in most of the radar profiles traversing the southern lake, accretion of lake water must be an active process in the southern half of the lake.

To examine the development and transport of the accreted ice, we have reconstructed an ice flow trajectory across the lake. Over the

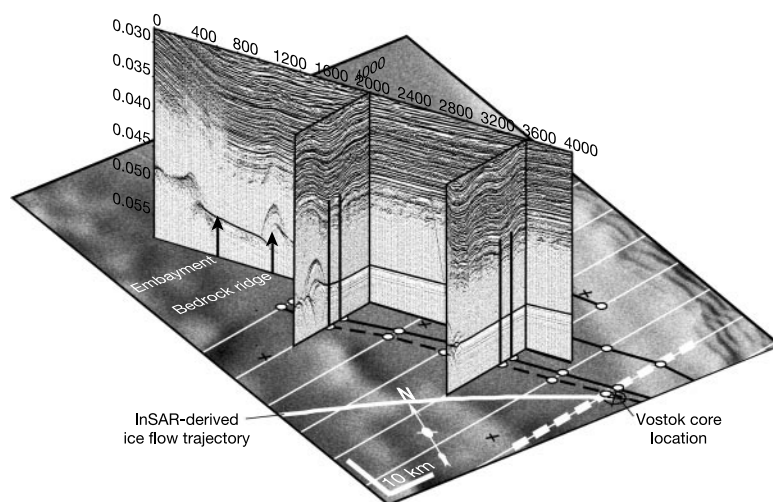


Figure 3 Fence diagram of the ice-penetrating radar data, illustrating the tracking of the internal layers across the lake superimposed on the RADARSAT image. Thin white lines show locations of ice-penetrating radar profiles used for analysis. Two east–west radar profiles are connected to a flow-parallel profile documenting the continuity of the internal layers from profile to profile. The east–west profiles illustrate the western shoreline and

the trajectory of the concave feature across the lake. The identification of the internal layer features are noted as white circles on the east–west lines. The thick black lines are the flow trajectories. The black dashed line is the flowline used to construct the ice thickness variations shown in Fig. 4. The white dashed line is the location of the Fig. 2 radar profile.

lake, the internal layers contain distinctive structures inherited from the topography on the western shore of the lake, such as the concave-down structure at the Vostok core site (Fig. 2). We follow this feature and three other distinctive internal layer structures in successive east–west profiles to construct trajectories across the lake (Fig. 1). Along the western rim of the lake, the trajectory is difficult to constrain as the shoreline is characterized by a 150–200-m bedrock ridge that separates a narrow embayment from the main lake (Fig. 1). We can track the inherited structures to the shoreline of the main lake at the bedrock ridge but not across the narrow embayment (Fig. 3). Our flow trajectories represent the ice flow field over Lake Vostok. The features were imprinted on the ice at the grounding line, and indicate that the flow over the lake was consistent for the time it has taken for the ice to traverse Lake Vostok. Our determination of the long-term flow field is supported by relatively flat internal layers of a radar line flown along the flow trajectories (Fig. 3). The only significant deformation of the internal layers along the flow-parallel profile occurs where ice flows into the lake at the western end of the line and over the bedrock ridge.

The long-term flow trajectories derived from tracing internal layers are parallel to the modern flow field measured with GPS positioning at six sites. The motion of the site at Vostok station, calculated using the base-station at McMurdo station, is $3.0 \pm 0.8 \text{ m yr}^{-1}$ at $131^\circ (\pm 4^\circ)$ relative to true north. Five other GPS stations located along a radar line (Fig. 2) show similar velocity vectors, parallel to the long-term flow field. Both the long-term flow trajectories derived from the radar data and the modern vectors based on GPS measurements differ significantly from the interferometric velocity vectors¹² determined by interferometric synthetic radar (InSAR) analysis (Fig. 1). Earlier estimates of melting and freezing⁴, based on the InSAR-derived velocity fields, are generally oblique to the ice flow and are contaminated by the inherited topography within the internal layers. Using the distance along the new flowlines from the western shoreline to the eastern shore (60 km) and from the bedrock ridge to the eastern shore (49 km), we can estimate the transit time for the ice sheet across the lake if the modern velocity (3.0 m yr^{-1}) has been constant. The shore-to-shore transit time is 20,000 years across the full width of the lake, and 16,000 years from the bedrock ridge to the eastern shore.

Along the Vostok core trajectory we have mapped five internal layers, the ice thickness and the accreted ice reflector (Fig. 3). Continuity of the internal layers from profile to profile is confirmed by linking the east–west profiles with the flow-parallel radar profile (Fig. 3). Along this flowline (Fig. 4), the ice sheet is 3,900 m thick when it enters the lake and thins 150 m over the lake in response to flow acceleration⁴ or divergence. The accreted ice layer nevertheless thickens, indicating continuing basal accretion. We have evidence for the accretion process both within the narrow embayment and over the main lake just to the east of the bedrock ridge. Along the flow-parallel line (Fig. 3), 60 m of accretion ice forms over a distance of 5 km within the embayment. This initial rapid thickening is consistent with simple thermal physics¹³, which predicts a thickening with time proportional to $(\kappa t)^{1/2}$, where κ is the thermal diffusivity of ice and t is the elapsed time since the onset of freezing. In this region, numerous diffractors, associated with bedrock peaks protruding into the ice sheet, basal crevasses or side reflection, characterize the lake reflector. The flow-parallel line traverses the embayment where the diffractors are present at the base of the ice sheet. In adjacent lines the bright lake reflector has significant topography, and is clearly disrupted by bedrock peaks; the embayment water is very shallow. As no accretion reflector is observed over the bedrock ridge, it remains unclear whether this ice is transported over the ridge or is sheared off into the shallow embayment. To the east of the bedrock ridge the accretion ice reflector is often imaged at the grounding line, diverging from the basal reflector. Along the Vostok trajectory, 120 m of accretion ice is first clearly imaged 16 km from the bedrock ridge. The accretion ice reaches a maximum thickness of 180 m at Vostok station, and thickens to 250 m as the ice sheet regrounds to the east of the lake.

We can estimate accretion rates along the flowline using the present-day ice velocity of 3 m yr^{-1} . In the narrow embayment an accretion rate of 3.8 cm yr^{-1} is required to produce the observed 60 m of accreted ice. The estimated accretion rate along the western shore will depend upon whether the ice accreted in the narrow embayment is transported over the bedrock ridge or remains within the embayment. If the accreted ice is transported over the ridge then the accretion rate over the western shore is 1.4 cm yr^{-1} , whereas if no embayment ice is transported over the main lake the accretion rate will be 2.5 cm yr^{-1} . In general, the accretion rate decreases along the

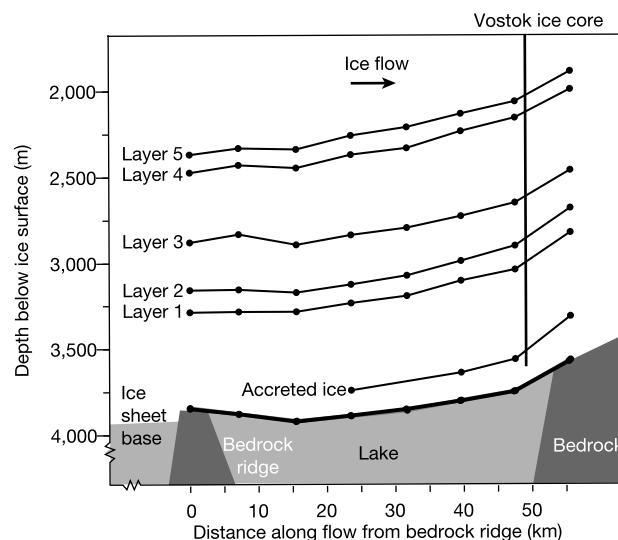


Figure 4 Ice thickness and accretion ice thicknesses along the Vostok flowline. The flowline was assembled from tracking the position of the base of the ice sheet, the accreted ice reflector, and five internal reflectors along the westernmost flow trajectory in

Figs 1 and 3 (black, dashed line). The western edge of the flow line is grounded over the bedrock ridge, and the eastern edge is grounded on the eastern edge of the lake (shown as cartoon).

flow line, reaching a minimum of 0.7 cm yr^{-1} at Vostok station. Along the eastern shoreline a significant increase in accretion rate to 2.9 cm yr^{-1} is necessary to produce the grounded accreted ice. It is possible that some of the accreted ice thickening along the eastern shoreline is the result of compressive flow and not accretion. The lower layers in the ice sheet (3,300–2,800 m) undergo a 7% thickening, in contrast to the shallower layers of the ice sheet (2,400–1,900 m) which thin by 5–13% when grounding. The accretion process is most evident along the lake shorelines.

In the grounded ice adjacent to the southeastern Lake Vostok shoreline, we observe an average of 295 m of accreted ice in 21 radar profiles (Fig. 1, inset). No accretion ice is imaged adjacent to the eastern grounding line in the northern 15 radar profiles. We estimate the ice flux out of the lake, projected onto a 165-km line downslope of the lake, parallel to local contours. Assuming an average thickness of 295 m of accretion ice and a mean velocity of 3 m yr^{-1} , we estimate an annual accreted ice volume flux of $0.146 \text{ km}^3 \text{ yr}^{-1}$ along this line. For a lake volume¹⁴ of $1,800 \text{ km}^3$, the residence time is 13,300 years. Previous estimates of residence time range from 4,500 years (ref. 6) to 125,000 years (ref. 1). The lower estimate (4,500 years), derived from the He^4/He^3 ratio⁶, may be related to the formation of the accreted ice samples along the western shoreline in regions isolated from the main lake circulation—that is, the shallow embayment or the western grounding line. Deeper samples of accreted ice should be more representative of the open lake.

In the southern portion of Lake Vostok, the interaction between the lake and the overlying ice sheet is dominated by accretion. Earlier evidence for melting in this region⁴ was based on radar data oblique to the ice flow field, and is, we believe, incorrect. Ice flow over southern Lake Vostok has a strong along-lake component, and has had a consistent orientation for the past 16,000 years—assuming that the velocity has been constant over this time period. The accretion process dominates at the lake shorelines, although some accretion continues over the middle of the lake. The accretion ice is generally transported out of the lake along the southeastern lake margin. Most samples of accretion ice analysed to date are derived either from the shallow embayment or from the western grounding line, environments that are not representative of the open lake system. □

Received 29 August 2001; accepted 11 February 2002.

- Kapitsa, A. P., Ridley, J. K., Robin, G. de Q., Siegert, M. J. & Zotikov, I. A. A large freshwater lake beneath the ice of central East Antarctica. *Nature* **381**, 684–686 (1996).
- Karl, D. M. *et al.* Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science* **286**, 2144–2147 (1999).
- Priscu, J. C. *et al.* Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science* **286**, 2141–2144 (1999).
- Siegert, M. J., Kwok, K., Mayer, C. & Hubbard, B. Water exchange between the subglacial Lake Vostok and the overlying ice sheet. *Nature* **403**, 643–646 (2000).
- Jouzel, J. *et al.* More than 200 meters of lake ice above subglacial Lake Vostok, Antarctica. *Science* **286**, 2138–2141 (1999).
- Jean-Baptiste, P., Petit, J.-R., Lipenkov, V. Y., Raynaud, D. & Barkov, N. I. Constraints on hydrothermal processes and water exchange in Lake Vostok from helium isotopes. *Nature* **411**, 460–462 (2001).
- Siegert, M. J. *et al.* Physical, chemical and biological processes in Lake Vostok and other Antarctic subglacial lakes. *Nature* **414**, 603–609 (2001).
- Popov, S. V., Mironov, A. V. & Sheremetiev, A. N. Average of the electromagnetic wave propagation velocity in ice measurements in Vostok station vicinity. *Materialy Glytsiologicheskikh Issledovaniy (Data of Glaciological Studies)* **90**, 206–208 (2001).
- Popkov, A. M., Kudryavtsev, G. A., Verkulich, S. R., Masolov, V. N. & Lukin, V. V. in *International Workshop on Lake Vostok Study: Scientific Objectives and Technological Requirements* 26 (Arctic and Antarctic Research Institute, St Petersburg, Russia, 1998).
- Whillans, I. M. Radio-echo layers and the recent stability of the West Antarctic ice sheet. *Nature* **264**, 152–155 (1976).
- Fujita, S. *et al.* Nature of radio echo layering in the Antarctic ice sheet detected by a two-frequency experiment. *J. Geophys. Res.* **B 104**, 13013–13024 (1999).
- Kwok, R., Siegert, M. J. & Carsey, F. D. Ice motion over Lake Vostok, Antarctica: constraints on inferences regarding the accreted ice. *J. Glaciol.* **46**, 689–694 (2000).
- Carlsaw, H. S. & Jaeger, J. C. *Conduction of Heat in Solids* 285 (Oxford Univ. Press, Oxford, 1959).
- Wüest, A. & Carmack, E. A priori estimates of mixing and circulation in the hard-to-reach water body of Lake Vostok. *Ocean Model.* **2**, 29–43 (2000).

15. Jezek, K., Noltimier, K. & The RAMP Product Team. *RAMP AMM-1 SAR Image Mosaic of Antarctica* [digital media] (Alaska SAR Facility, Fairbanks; and the National Snow and Ice Data Center, Boulder, Colorado, 2001).

Acknowledgements

The ice-penetrating radar data were acquired by the US National Science Foundation's Support Office for Aerogeophysical Research (SOAR) located at the University of Texas. We acknowledge the contributions of the Vostok field team, including the SOAR team, the Kenn Borek flight crews and the Raytheon East Camp crew. Comments from C. Bentley, R. Alley, E. Waddington and M. Siegert were appreciated. RADARSAT satellite data was provided by the Canadian Space Agency. This work was supported by the US National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.E.B. (e-mail: robinb@ldeo.columbia.edu).

Development of anisotropic structure in the Earth's lower mantle by solid-state convection

Allen K. McNamara*, Peter E. van Keken* & Shun-ichiro Karato†

* Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109-1063, USA

† Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520-8109, USA

Seismological observations reveal highly anisotropic patches at the bottom of the Earth's lower mantle, whereas the bulk of the mantle has been observed to be largely isotropic^{1–4}. These patches have been interpreted to correspond to areas where subduction has taken place in the past or to areas where mantle plumes are upwelling, but the underlying cause for the anisotropy is unknown—both shape-preferred orientation of elastically heterogeneous materials⁵ and lattice-preferred orientation of a homogeneous material^{6–8} have been proposed. Both of these mechanisms imply that large-strain deformation occurs within the anisotropic regions, but the geodynamic implications of the mechanisms differ. Shape-preferred orientation would imply the presence of large elastic (and hence chemical) heterogeneity whereas lattice-preferred orientation requires deformation at high stresses. Here we show, on the basis of numerical modelling incorporating mineral physics of elasticity and development of lattice-preferred orientation, that slab deformation in the deep lower mantle can account for the presence of strong anisotropy in the circum-Pacific region. In this model—where development of the mineral fabric (the alignment of mineral grains) is caused solely by solid-state deformation of chemically homogeneous mantle material—anisotropy is caused by large-strain deformation at high stresses, due to the collision of subducted slabs with the core–mantle boundary.

Lattice-preferred orientation (LPO) leads to the development of a mineral fabric in material that deforms primarily by dislocation creep. Previous work⁹ has shown that slabs cause high-stress regions in the lower mantle, which leads to localized regions of dislocation creep within a lower mantle dominated by diffusion creep under a wide range of rheological parameters. It has also been shown that hot, upwelling regions are low stress, and are therefore dominated by diffusion creep. The presence of dislocation creep is not in itself a

sufficient condition for the formation of LPO in slabs. In addition, the strain due to the dislocation creep must be greater than 100–200%. Here we investigate the circumstances under which this condition is met by a combined mineral-physics and dynamical-modelling approach. This requires the use of a composite rheology formulation involving a combination of diffusion and dislocation creep deformation mechanisms.

We choose rheological parameters for the upper and lower mantle based on mineral-physics observations^{10,11}, resulting in a viscosity profile that generally increases with depth and includes an increase at the base of the transition zone. We also use a yield stress approach¹² in the upper regions of our model to form subducting slabs. We place strain tracers in slab regions above the transition zone to track the evolution of deformation. Given the uncertainties related to first-principles calculations of fabric development, such as the critical shear stresses associated with specific slip systems and the nature of dynamic recrystallization, we assume that strain may be used as a proxy for the development of mineral fabric. During dislocation creep, material flows by the slipping of specific glide planes, resulting in an oriented array of crystal axes and promoting fabric development. On the other hand diffusion creep occurs by the migration of atoms involving grain-boundary sliding, resulting in a random orientation of crystal axes and tending to destroy any pre-existing fabric. Therefore, we track strain only in regions dominated by dislocation creep. We compare the resulting strain field to mineral-physics deformation experiments in order to assess the expected seismic anisotropy.

The numerical calculations of mantle convection are solved using the conservation equations of mass, momentum and energy in the extended Boussinesq formulation¹³ using a finite-element approach. The modelling geometry is a two-dimensional quarter-cylinder with free-slip boundaries. We employ depth-dependent thermal expansivity and thermal conductivity^{14,15}. We allow our convection calculations to run for long enough to ensure that the initial condition does not affect the results. Details of the model setup and rheological formulation are given in previous work⁹. Parameters used for the calculations presented here are given in Table 1.

Lagrangian finite strain is calculated as a post-processing step,

using the velocity fields at each time step of the convection calculation. Strain is calculated for particles within regions dominated by dislocation creep by time-integrating the deformation gradient tensor for each strain tracer¹⁶. The strain calculation code was checked by comparison to analytical solutions of simple and pure shear as well as a combination of the two¹⁷. If a strain tracer leaves the regime of dislocation creep, its recorded strain magnitude is decreased as a function of further material strain. We assume that LPO is destroyed after the material stretches further than twice its original length.

The results of two calculations are given here (Figs 1 and 2) to illustrate the insensitivity of our results to the amount of internal heating and to the scaling geometry. The calculations vary in terms of heating mode and radial scaling. Figure 1 shows results from a calculation in which the mantle is entirely bottom-heated and scaled to preserve the volume ratio of the Earth's upper and lower mantle. Figure 2 shows results from a calculation in which the mantle is heated by both internal and bottom heating (~50% each), and is scaled to preserve the surface area to volume ratio of the mantle in a spherical Earth. These scalings have been found to better approximate the heat and mass transfer of spherical models¹⁸. Both results yield heat flow and radial viscosity profiles that are consistent with observations^{19,20}.

Snapshots in time of the temperature and viscosity ratio fields are shown in Figs 1c–e and 2c–e. The viscosity ratio is defined as the ratio of the components corresponding to dislocation creep and diffusion creep. Regions in blue represent a positive ratio, indicating that they are dominated by diffusion creep. Red regions represent a negative ratio, indicating a domination of dislocation creep. Note that most of the upper mantle flows by dislocation creep. In contrast the interior of the upper-mantle slab flows primarily by diffusion creep, because the upper-mantle activation coefficients result in a transition stress that increases with decreasing temperature. Away from the slab region, lower-mantle flow is dominated by diffusion creep, particularly in the lowermost boundary region, where high temperatures lead to low viscosities and therefore low stresses.

Strain tracers are shown as points on the viscosity ratio fields. Superimposed on tracer points are vectors representing the finite

Table 1 Calculation parameters

Parameter	Description	Units	Value, Fig. 1	Value, Fig. 2
ΔT	Temperature drop across mantle	K	3,000	3,000
α_0	Reference thermal expansivity	K ⁻¹	3×10^{-5}	3×10^{-5}
ρ_0	Reference density	kg m ⁻³	4,500	4,500
C_p	Specific heat	J kg ⁻¹ K ⁻¹	1,250	1,250
h	Mantle thickness	m	2.8×10^6	2.8×10^6
k_0	Reference thermal conductivity	W m ⁻¹ K ⁻¹	5.6	5.6
g	Gravitational constant	m s ⁻²	9.8	9.8
κ_0	Reference thermal diffusivity	m ² s ⁻¹	10^{-6}	10^{-6}
Di	Dissipation number		0.5	0.5
d_{um}	Upper mantle grain size	mm	2.0	2.0
d_{lm}	Lower mantle grain size	mm	1.0	1.0
m	Grain size index		2.5	2.5
n	Power law index		3.0	3.0
$A_{diff-um+}$	Upper-mantle diffusion creep prefactor	d _{um} ^m Pa ⁻¹ s ^{-1*}	1.3276×10^{-13}	2.6551×10^{-14}
$A_{diff-lm+}$	Lower-mantle diffusion creep prefactor	d _{lm} ^m Pa ⁻¹ s ^{-1*}	6.2230×10^{-14}	1.3276×10^{-13}
$A_{disl-um+}$	Upper-mantle dislocation creep prefactor	Pa ⁿ s ^{-1*}	4.7295×10^{-13}	9.4590×10^{-14}
$A_{disl-lm+}$	Lower-mantle dislocation creep prefactor	Pa ⁿ s ^{-1*}	1.2446×10^{-29}	2.6551×10^{-29}
$g_{diff-um}$	Upper-mantle diffusion creep activation coefficient		17	17
$g_{diff-lm}$	Lower-mantle diffusion creep activation coefficient		10	10
$g_{disl-um}$	Upper-mantle dislocation creep activation coefficient		31	31
$g_{disl-lm}$	Lower-mantle dislocation creep activation coefficient		10	10
σ_d	Ductile yield stress	MPa	400	250
σ_b	Brittle yield stress gradient	MPa km ⁻¹	5.33	3.33
$R_{surface}$	Non-dimensional surface radius		1.67813	1.4292
R_{bottom}	Non-dimensional bottom radius		0.67813	0.4292
$R_{interface}$	Non-dimensional upper—lower mantle boundary radius		1.42	1.19
η_{max}	Non-dimensional maximum viscosity		1.0	1.0
η_{min}	Non-dimensional minimum viscosity		10^{-6}	10^{-6}

* Indices m and n are the grain size index and the power law index, respectively.

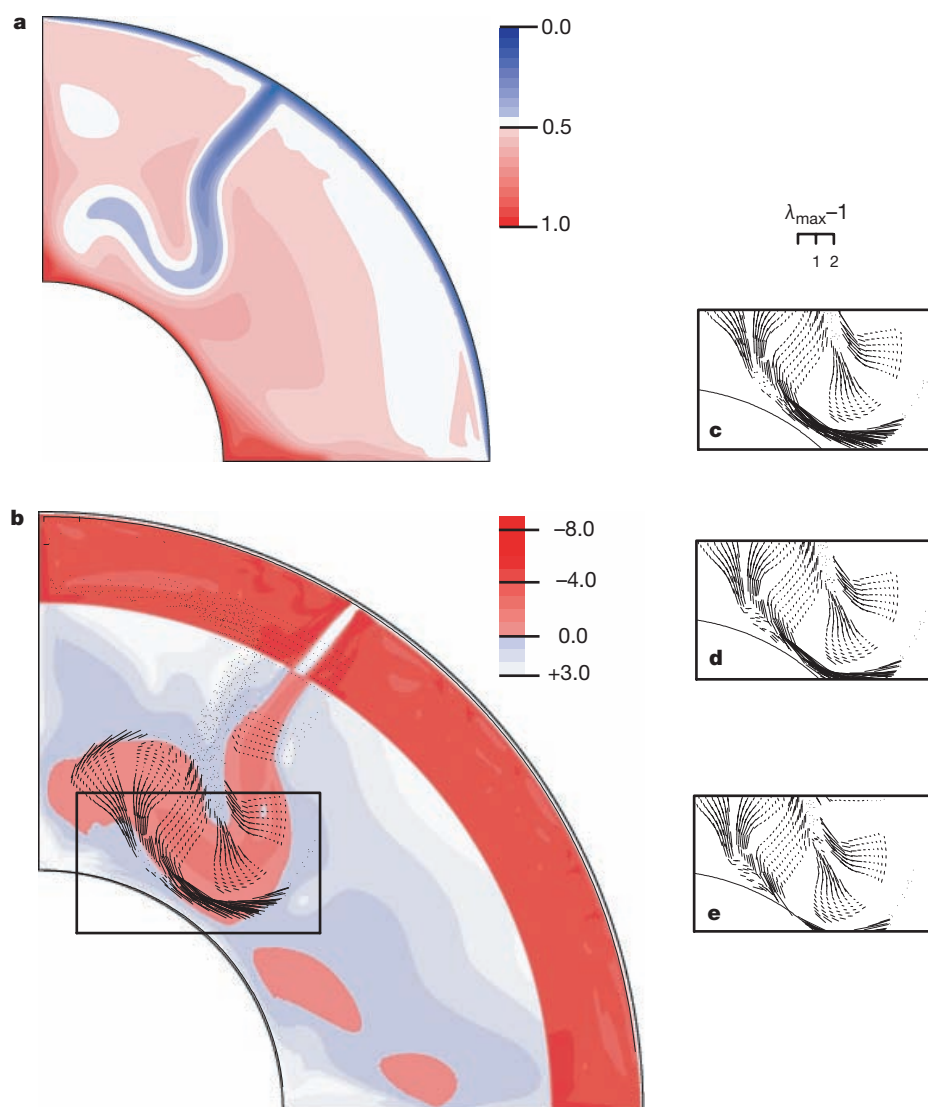


Figure 1 Snapshot of a slab impinging on the core-mantle boundary (CMB) in the bottom-heated model. **a**, Non-dimensional temperature field and **b**, ratio of viscosity due to dislocation creep over that due to diffusion creep. Positive values indicate deformation is dominated by diffusion creep, and negative values indicate deformation is dominated by

dislocation creep. Superimposed are strain tracers and their associated strain vectors. Strain vectors represent maximum stretch, $\lambda_{\max}$, and have a magnitude proportional to the stretch minus 1. The evolution of strain of the area of **b** shown in a box is given after successive intervals of 25 Myr in **c**, **d**, **e**.

strain. The vectors are in the direction of maximum principle strain, and have a length related to the stretch. The stretch is defined as the ratio of final length to the original length. Therefore, an undeformed particle has a stretch of unity. To differentiate better between strained and unstrained tracers the vector length is set to the stretch minus one, so only strained tracers have an associated vector.

To illustrate the temporal evolution of strain, Figs 1c–e and 2c–e show the strain configurations at successive times. Results from Fig. 1 reveal that although slab deformation in the lower mantle is dominated entirely by dislocation creep, the strain is not well developed because of low strain rates in the viscous slab. Directly above the core-mantle boundary (CMB), however, the magnitude of strain increases dramatically, resulting in a region of high strain that is directed laterally. Figure 1c–e reveals the time-dependent nature of the strain configuration. The overall strain magnitude in Fig. 1e is greatly reduced, and the directional sense is more random. Results are similar for the calculation

shown in Fig. 2. Again, deformation in the slab is dominated by dislocation creep in a mantle otherwise dominated by diffusion creep. Figure 2b reveals a low magnitude lateral sense of strain directly above the CMB that increases with time (Fig. 2c–e) resulting in a strong laterally directed strain field. The strain markers also show that material is rotated as it approaches the upwellings that are constrained to the side boundaries of the model.

Our results show that the details of the strain field are time-dependent and can be quite complicated. Analysis of numerous convection results reveal that a lateral sense of high-magnitude extension (in excess of 100%) directly above the CMB is a general feature of the strain field. This feature tends to be long-lived but not permanent. As shown here, the strain pattern may decay into one that is characterized by a lower magnitude and inconsistent direction (that is, the progression from Fig. 1b–e). We also see that a coherent pattern of high-magnitude lateral strain tends to develop from a less developed state (that is, the progression

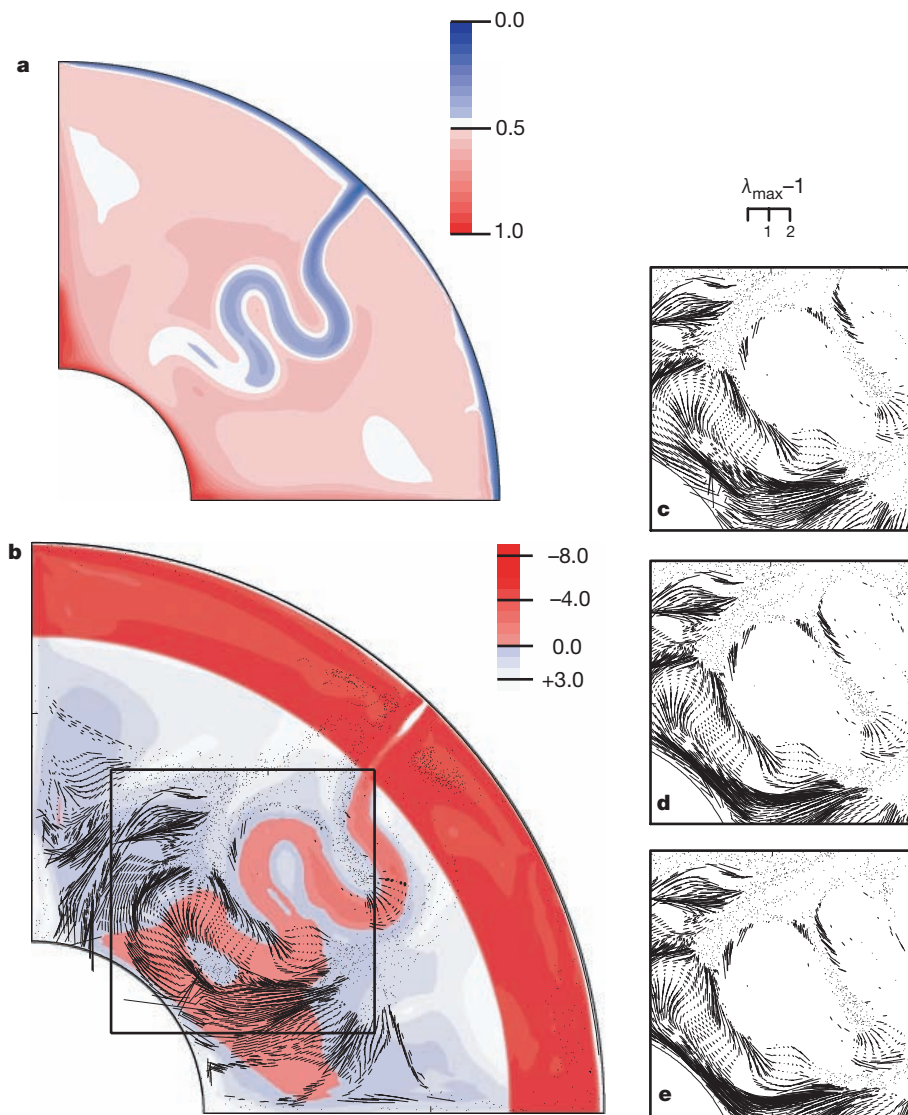


Figure 2 As Fig. 1, but for a model with bottom heating and internal heating.

from Fig. 2b–e). In general, we find that slab regions are characterized by a high degree of lateral strain directly above the CMB that occasionally reverts to a low-magnitude, more randomly oriented strain.

In order to assess the implications for seismic anisotropy, comparison to mineral-physics data is required. Experimental studies on LPO development show that the strength of LPO increases with strain, and reaches nearly steady-state values at a certain strain^{21–23}. This critical strain depends on the material as well as deformation conditions. A detailed experimental study of LPO development is now available for (Mg,Fe)O (ref. 23). Because of its large elastic anisotropy²⁴, LPO of (Mg,Fe)O results in a detectable seismic anisotropy (~1–2%)⁵, although its volume fraction is small (less than 20%). For (Mg,Fe)O, a shear strain of ~200% is needed to develop significant LPO and steady-state LPO is formed at strains of ~400–500% (ref. 23). At steady state, LPO of (Mg,Fe)O corresponding to horizontal shear results in $V_{SH} > V_{SV}$ anisotropy²³ where V_{SH} and V_{SV} correspond to horizontal and vertical components of the shear wave velocity, respectively.

Our numerical models show the development of strong subhorizontal shear strain near the base of the mantle. Therefore, LPO of

(Mg,Fe)O provides a natural explanation for the spatial variation in anisotropy in the lower mantle. Contributions from another major component, (Mg,Fe)SiO₃ perovskite, are difficult to estimate because of the absence of experimental data on LPO. In the two-phase mixture, (Mg,Fe)O is probably the weaker phase¹⁰ and deforms more easily than (Mg,Fe)SiO₃. As a consequence, a higher degree of strain will occur in (Mg,Fe)O for a given deformation. In addition, if the results of LPO development in analogous material are used, the contribution from (Mg,Fe)SiO₃ perovskite is likely to be smaller than that of (Mg,Fe)O and the sense is opposite ($V_{SV} > V_{SH}$ for horizontal shear)^{6,23}. Once materials leave high-stress regions, diffusion creep dominates and erases pre-existing LPO. Our models show highly time-dependent flow patterns, so the regions of strong anisotropy in these models correspond to regions with high-stress, large-strain deformation in the recent past. In contrast, if shape-preferred orientation due to laminated structures of palaeo-crust and ambient mantle is responsible for the anisotropy⁵, then anisotropic structures will have much longer lifetimes and would not represent recent dynamic regimes. Also, the origin of the high contrast in elastic properties necessary for shape-preferred orientation is not clear. (We note that the origin of anisotropy in the Central Pacific, probably related to plume upwell-

ing, is not well constrained by the present work. It is possible that anisotropy in these regions is caused by shape-preferred orientation involving aligned melt pockets.)

Our results show that slabs are characterized by high stress, resulting in deformation dominated by dislocation creep within a mantle otherwise dominated by diffusion creep. We find complicated strain fields associated with deformation due to dislocation creep, but one consistent feature appears to be a large degree of laterally directed strain directly above the CMB. When examined in the context of mineral-physics experiments, we predict that this strain field results in significant seismic anisotropy, with $V_{SH} > V_{SV}$. This work shows that LPO of (Mg, Fe)O is a likely candidate for the seismic anisotropy observed near the CMB in slab regions. Although other processes may contribute to the formation of anisotropy⁵, they are not required, and solid-state processes within a homogeneous material may suffice. □

Received 24 October 2001; accepted 31 January 2002.

1. Lay, T., Williams, Q. & Garnero, E. J. The core–mantle boundary layer and deep Earth dynamics. *Nature* **392**, 461–468 (1998).
2. Lay, T., Williams, Q., Garnero, E. J., Kellogg, L. & Wyssession, M. E. in *The Core–Mantle Boundary* (eds Gurnis, M., Wyssession, M. E., Knittle, E. & Buffett, B. A.) 299–318 (Geodynamics Series Vol. 28, Am. Geophys. Union, Washington DC, 1998).
3. Kendall, J. M. in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale* (eds Karato, S., Forte, A. M., Liebermann, R. C., Masters, G. & Stixrude, L.) 133–159 (Geophysics Monograph 117, Am. Geophys. Union, Washington DC, 2000).
4. Ritsema, J. Evidence for shear velocity anisotropy in the lowermost mantle beneath the Indian Ocean. *Geophys. Res. Lett.* **27**, 1041–1044 (2000).
5. Kendall, J. M. & Silver, P. G. Constraints from seismic anisotropy on the nature of the lowermost mantle. *Nature* **381**, 409–412 (1996).
6. Karato, S. Some remarks on the origin of seismic anisotropy in the D'' layer. *Earth Planets Space* **50**, 1019–1028 (1998).
7. Karato, S. Seismic anisotropy in the deep mantle, boundary layers and the geometry of mantle convection. *Pure Appl. Geophys.* **151**, 565–587 (1998).
8. Stixrude, L. in *The Core–Mantle Boundary* (eds Gurnis, M., Wyssession, M. E., Knittle, E. & Buffett, B. A.) 83–96 (Geodynamics Series Vol. 28, Am. Geophys. Union, Washington DC, 1998).
9. McNamara, A. K., Karato, S. & van Keken, P. E. Localization of dislocation creep in the lower mantle: Implications for the origin of seismic anisotropy. *Earth Planet. Sci. Lett.* **191**, 85–99 (2001).
10. Yamazaki, D. & Karato, S. Some mineral physics constraints on the rheology and geothermal structure of Earth's lower mantle. *Am. Mineral.* **86**, 385–391 (2001).
11. Frost, H. J. & Ashby, M. F. *Deformation Mechanism Maps* (Pergamon, Oxford, 1982).
12. Tackley, P. J. Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations: 1. Pseudoplastic yielding. *Geochim. Geophys. Geosys.* [online] (<http://g-cubed.org/>) **1**, 2000GC000036 (2000).
13. Ita, J. & King, S. D. Sensitivity of convection with an endothermic phase change to the form of governing equations, initial conditions, boundary conditions, and equations of state. *J. Geophys. Res.* **99**, 15919–15938 (1994).
14. Jarvis, G. T. & McKenzie, D. P. Convection in a compressible fluid with infinite Prandtl number. *J. Fluid Mech.* **96**, 515–583 (1980).
15. Leitch, A. M., Yuen, D. A. & Sewell, G. Mantle convection with internal heating and pressure-dependent thermal expansivity. *Earth Planet. Sci. Lett.* **102**, 213–232 (1991).
16. Spencer, A. J. M. *Continuum Mechanics* (Longman, London, 1980).
17. Ramberg, H. Particle paths, displacement and progressive strain applicable to rocks. *Tectonophysics* **28**, 1–37 (1975).
18. van Keken, P. E. Cylindrical scaling for dynamical cooling models of the Earth. *Phys. Earth Planet. Inter.* **124**, 119–130 (2001).
19. Mitrova, J. X. & Forte, A. M. Radial profile of mantle viscosity: results from the joint inversion of convection and postglacial rebound observables. *J. Geophys. Res.* **102**, 2751–2769 (1997).
20. Pollack, H. N., Hurter, S. J. & Johnson, J. R. Heat flow from the Earth's interior: analysis of the global data set. *Rev. Geophys.* **31**, 267–280 (1993).
21. Zhang, S. & Karato, S. Preferred orientation of olivine deformed in simple shear. *Nature* **375**, 774–777 (1995).
22. Mainprice, D., Barroul, G. & Ben Ismail, W. in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale* (eds Karato, S. et al.) 237–264 (Am. Geophys. Union, Washington DC, 2000).
23. Yamazaki, D. & Karato, S. Fabric development in (Mg,Fe)O during large strain, shear deformation: Implications for seismic anisotropy in Earth's lower mantle. *Phys. Earth Planet. Inter.* (in the press).
24. Karki, B. B. et al. Structure and elasticity of MgO at high pressure. *Am. Mineral.* **82**, 51–60 (1997).

Acknowledgements

We thank M. Gurnis, D. Yamazaki, R. Wenk, T. Lay and B. Kiefer for discussions. This research was supported by the CSEDI program of the National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.K.M. (e-mail: mcnamar@umich.edu).

A ceratopsian dinosaur from China and the early evolution of Ceratopsia

Xing Xu*, Peter J. Makovicky†, Xiao-lin Wang*, Mark A. Norell‡ & Hai-lu You*

* Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, 100044, China

† The Field Museum, 1400 S. Lake Shore Drive, Chicago, Illinois 60605, USA

‡ American Museum of Natural History, Central Park West at 79th St., New York 10024, USA

Ceratopsians (horned dinosaurs) represent one of the last and the most diverse radiations of non-avian dinosaurs^{1–4}. Although recent systematic work unanimously supports a basal division of Ceratopsia into parrot-like psittacosaurids and frilled neoceratopsians, the early evolution of the group remains poorly understood, mainly owing to its incomplete early fossil record. Here we describe a primitive ceratopsian from China. Cladistic analysis posits this new species as the most basal neoceratopsian. This new taxon demonstrates that some neoceratopsian characters evolved in a more incremental fashion than previously known and also implies mosaic evolution of characters early in ceratopsian history.

The lacustrine lower part of the Yixian Formation of western Liaoning, China, has produced many spectacular fossil remains, including feathered dinosaurs⁵. Recently a number of important three-dimensionally preserved vertebrate fossils, including specimens of a new ceratopsian, have been collected from the fluvial facies that form the lowest part of the Yixian Formation⁶. Some radiometric analyses suggest these deposits are older than 128 Myr and younger than 139 Myr (ref. 5) (Early Cretaceous period) but others imply they are older than 145 Myr (ref. 7) (Late Jurassic period).

Ceratopsia Marsh, 1890

Neoceratopsia Sereno, 1986

Liaoceratops yanzigouensis gen. et sp. nov.

Etymology. The generic name is derived from the provincial name 'Liaoning', and the suffix 'ceratops', commonly used for horned dinosaur names. The specific name 'yanzigou' refers to the village near which the holotype was found.

Diagnosis. Neoceratopsian characterized by sutures between premaxilla, maxilla, nasal and prefrontal intersecting at a common point high on the side of the snout, possession of several tubercles on the ventral margin of the angular, a foramen on the posterior face of the quadrate near the articulation with the quadratojugal, a small tubercle on the dorsal border of the foramen magnum, and a thick posterior border of the parietal frill.

Holotype. IVPP (Institute of Vertebrate Paleontology and Paleoanthropology, Beijing) V12738, an almost complete skull (Fig. 1).

Referred specimen. IVPP V12633, a juvenile skull (Fig. 2).

Locality and horizon. Yanzigou and Lujiatun, Shangyuan, western Liaoning, China; the lowest part of the Yixian Formation, probably Early Cretaceous.

Description. The holotype skull of *Liaoceratops* is comparable in size to other basal ceratopsians such as *Psittacosaurus* and *Chaoyangsaurus*. Progressive, but incomplete, sutural closures between skull elements suggest that it derives from a subadult individual. The rostral appears unkeeled as in *Psittacosaurus*. As in many neoceratopsians, the rostral bears lateral processes along the buccal margin. The premaxilla forms much of the side of the snout and extends posterodorsally to reach the prefrontal, as in psittacosaurids and basal ornithomorphs. Three cylindrical premaxillary teeth are present as in the primitive neoceratopsian *Archaeoceratops*⁸. The snout is relatively wide in dorsal view, and tapers abruptly, being intermediate in form between that of psitta-

cosaur and neoceratopsians. A large triangular antorbital fossa is present as in *Archaeoceratops*⁸. The jugal bears a low, lateral horn, which is positioned anterior to the quadrate rather than posterior to it as in higher neoceratopsians. A separate epijugal is apparently absent in *Liaoceratops*. A long posterodorsal process of the jugal reaches the squamosal and forms most of the anterior and dorsal border of the infratemporal fenestra. The quadratojugal is exposed in lateral view as in psittacosaur, but unlike in neoceratopsians—in which it is covered by the posteriorly expanded jugal horn. The skull roof is flat, and the fused parietals bear a tall sagittal crest as in *Protoceratops*⁹. *Liaoceratops* possesses a short parietal frill, which has a rounded posterior border as in *Leptoceratops*. The squamosal contributes to the lateral part of the frill. Irregular, asymmetric openings are present on each side of the frill midline, but these are, at least in part, preservational artefacts and presence or absence of frill fenestration cannot be established with certainty. The parietal is very thin rostral to the posterior margin, which is exceptionally robust. The rostral face of this thick bar is densely pitted indicating muscular insertion.

The buccal cavity is expanded at the level of the premaxillary palate, and the buccal margin of the maxilla–premaxilla forms a sinuous curve in ventral view. The tooth row is short, and not expanded behind the orbit as in higher neoceratopsians. The postpalatal part of the pterygoid is long and bears a large posterior process as in some psittacosaur and several basal neoceratopsian taxa such as *Leptoceratops*¹⁰. The paired pterygoid processes cover most of the basicranium from view. The ectopterygoid is large and extends from its contact with the jugal to the tip of the elongate mandibular process of the pterygoid. The occipital condyle is spherical, but the basioccipital may participate in the foramen magnum. A narrow ridge extends along the basioccipital midline, ventral to the occipital condyle as in *Archaeoceratops*⁸ and *Microceratops*¹¹.

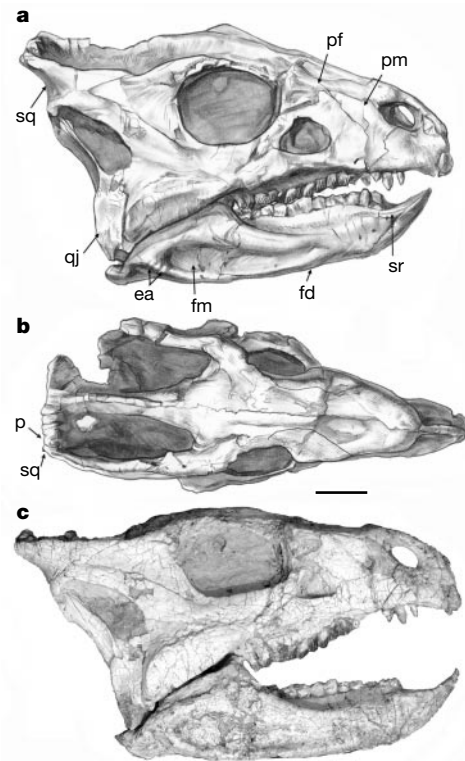


Figure 2 Skull of the juvenile specimen referred to *Liaoceratops yanzigouensis* (IVPP V12633) in lateral (**a**, **c**) and dorsal (**b**) views. Scale bar, 1 cm. Abbreviations as in Fig. 1, except for Sr, sharp ridge.

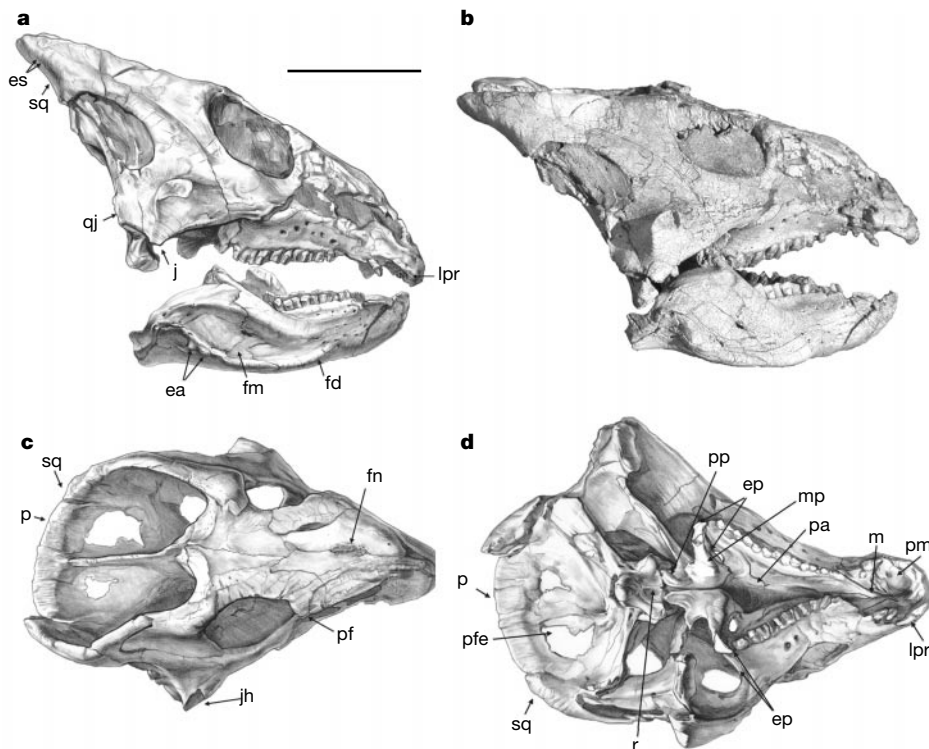


Figure 1 Holotype skull of the ceratopsian *Liaoceratops yanzigouensis* (IVPP V12738) in lateral (**a**, **b**), dorsal (**c**) and ventral (**d**) views. Scale bar, 4.5 cm. Abbreviations: ea, eminences on angular; ep, ectopterygoid; es, eminences on squamosal; fd, flange on dentary; fm, fossa on external surface of the mandible; fn, fossa on nasals; j, jugal; jh,

jugal horn; lpr, lateral process of jugal; m, maxilla; mp, mandibular process of pterygoid; p, parietal; pa, palatine; pf, prefrontal; pm, premaxilla; pp, posterior process of pterygoid; r, ridge; pfe, parietal fenestra; qj, quadratojugal; sq, squamosal.

The prementary is hooked dorsally and has a bevelled edge as in basal neoceratopsians. The dentary is massive and bears a small ventral flange. As in neoceratopsians, the surangular bulges laterally, but it lacks the lateral wall to the glenoid present in other members of the group. The retroarticular process is short.

Weak, incipient primary ridges are present on the cheek teeth of the holotype, and are less pronounced in the referred juvenile specimen. Where visible, the roots appear to lack lateral grooves, and the eruption pattern in the dentary is not as rigidly determinate as in neoceratopsians although more ordered than in psittacosaurids and *Chaoyangsaurus*¹². Cingula are not developed near the crown bases, and worn teeth display oblique wear facets as in primitive ornithischians.

The referred juvenile skull is about half as long as the holotype. It differs from the holotype of *Liaoceratops* in features characteristic of juvenile ceratopsians^{9,13,14}, including fewer teeth, vaulted frontals, a weaker jugal horn and a proportionately shorter and narrower frill. Similar differences are observed in the growth series of other basal ceratopsians such as *Psittacosaurus*¹⁵ and *Protoceratops*⁹. Comparison between the juvenile and holotype specimens shows that the squamosal contribution to the frill margin increases allometrically with growth and the squamosals form half the frill margin in the adult. This contrasts with *Protoceratops*, in which the squamosal contribution to the frill decreases proportionately throughout ontogeny.

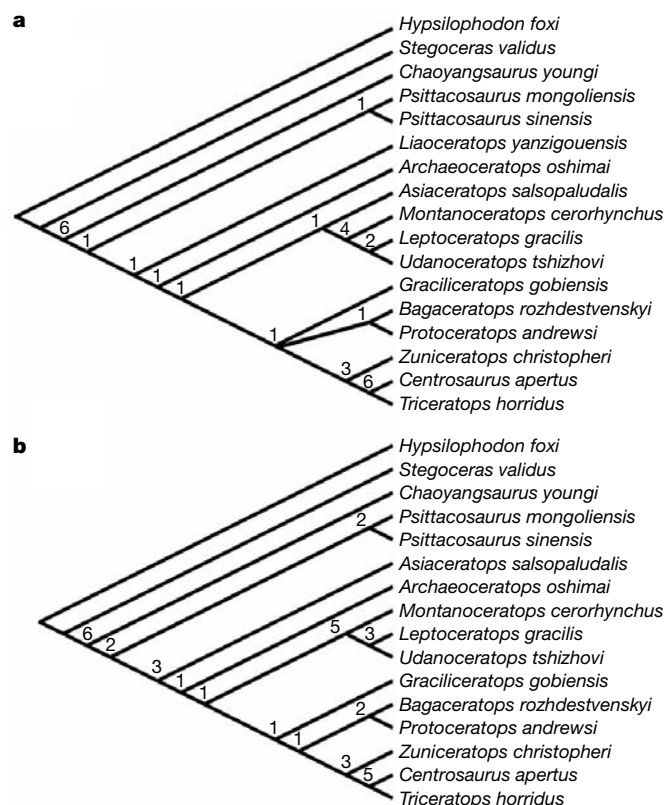


Figure 3 Phylogenetic relationships among ceratopsians. **a**, A strict consensus cladogram of two most-parsimonious trees (200 steps; consistency index, 0.69; retention index, 0.78) derived from analysis of our data set (based in part on refs 4, 17, 18) placing *Liaoceratops* as the most basal neoceratopsian. Integers above internodes are decay indices. The strict and Adams consensus trees have identical topologies. **b**, Single most-parsimonious cladogram (190 steps; consistency index, 0.73; retention index, 0.81) based on phylogenetic analysis of the same data set with *Liaoceratops yanzigouensis* excluded. We note the different phylogenetic positions of *Archaeoceratops* and *Asiaceratops* and higher decay indices relative to **a** after exclusion of *Liaoceratops yanzigouensis*.

Discussion. Cladistic analysis (see Supplementary Information) posits *Liaoceratops* as the most basal neoceratopsian (Fig. 3a). Derived characters shared with other neoceratopsians include the lateral processes of the rostral, an expanded frill with squamosal participation, a spherical occipital condyle, a deep temporal bar, a triangular postorbital and laterally convex surangular. *Liaoceratops* still retains a number of primitive ceratopsian characters, however. For example, the quadratojugal is flat rather than transversely expanded, the rostral is unkeeled, an epijugal is absent, and the maxillary teeth have weakly developed primary ridges and oblique occlusion angles. Indeed, some character states in *Liaoceratops* are intermediate between those of psittacosaurids and higher neoceratopsians, documenting an incremental evolution for certain neoceratopsian diagnostic characters.

We note that *Liaoceratops* also exhibits characters traditionally used to diagnose either psittacosaurids or more exclusive clades within Neoceratopsia. For example, *Liaoceratops* bears a weak ventral flange on the dentary and has an infratemporal fenestra that is wider ventrally, as in some, but not all, *Psittacosaurus* species¹⁶. Analysis of the data set with *Liaoceratops* excluded results in a decrease in tree length of ten steps, and a concomitant increase in the support for nodes near the base of Ceratopsia, including Psittacosauridae, Neoceratopsia and the clade that unites them (Fig. 3b). The mosaic character distribution in *Liaoceratops* introduces homoplasy to the data, which strips away some traditional psittacosaurid synapomorphies, and repolarizes certain characters within Neoceratopsia, causing a reversal in the phylogenetic order between *Asiaceratops* and *Archaeoceratops*.

Liaoceratops is the oldest known neoceratopsian, and represents a significant stratigraphic range extension for this clade, whose other members are Albian⁸ or younger. *Liaoceratops* co-occurs with specimens of *Psittacosaurus* in the lowermost part of the Yixian Formation, whereas *Chaoyangsaurus* is known from the underlying, probably Late Jurassic¹² Tuchengzi Formation. The basal ceratopsian split between psittacosaurids and neoceratopsians occurred no later than the earliest part of the Cretaceous, and both lineages appear to have acquired some of their diagnostic features rapidly within the latest part of the Jurassic and possibly the earliest part of the Cretaceous. The combination of these temporal constraints with the previously unsuspected mosaic evolution introduced by *Liaoceratops* indicates a more rapid rate of character evolution at the base of Ceratopsia and its major subclades than was hitherto recognized⁴.

Received 2 October; accepted 17 December 2001.

1. Dodson, P. & Currie, P. In *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 593–618 (Univ. California Press, Berkeley, 1990).
2. Dodson, P. In *Encyclopedia of Dinosaurs* (eds Currie, P. J. & Padian, K.) 473–478 (Academic, San Diego, 1997).
3. Sereno, P. The origin and evolution of dinosaurs. *Annu. Rev. Earth Planet. Sci.* **25**, 234–256 (1997).
4. Sereno, P. C. In *The Age of Dinosaurs in Russia and Mongolia* (eds Benton, M. J., Shishkin, M. A., Unwin, D. M. & Kurochkin, E. N.) 480–516 (Cambridge Univ. Press, New York, 2000).
5. Swisher, C. C. et al. Further support for a Cretaceous age for the feathered-dinosaur beds of Liaoning, China: new ⁴⁰Ar/³⁹Ar dating of the Yixian and Tuchengzi Formations. *Chinese Sci. Bull.* **46**, 2009–2013 (2001).
6. Xu, X., Wang, X.-L. & You, H.-L. A primitive ornithomimid from the Early Cretaceous Yixian Formation of Liaoning. *Vertebrata Palasiatica* **38**, 318–325 (2000).
7. Lo, C.-H., Chen, P.-J., Tsou, T.-Y., Sun, S.-S. & Lee, C. Y. ⁴⁰Ar/³⁹Ar laser single-grain and K–Ar dating of the Yixian Formation, NE China. *Palaeoworld* **11**, 328–340 (1999).
8. Dong, Z.-M. & Azuma, Y. In *Sino-Japanese Silk Road Dinosaur Expedition* (ed. Dong, Z.-M.) 68–89 (China Ocean, Beijing, 1997).
9. Brown, B. & Schlaikjer, E. M. The structure and relationships of *Protoceratops*. *Ann. NY Acad. Sci.* **40**, 133–266 (1940).
10. Sternberg, C. M. Complete skeleton of *Leptoceratops gracilis* Brown from the Upper Edmonton Member on Red Deer River, Alberta. *Nat. Mus. Can. Bull.* **123**, 225–255 (1975).
11. Bohlin, B. *Fossil Reptiles from Mongolia and Kansu* (The Sino-Swedish Expedition, Statens Etnografiska Museum, Stockholm, 1953).
12. Zhao, X.-J., Cheng, Z.-W. & Xu, X. The earliest ceratopsian from the Tuchengzi Formation of Liaoning, China. *J. Vert. Paleontol.* **19**, 681–691 (1999).
13. Maryanska, T. & Osmolska, H. Protoceratopsidae (Dinosauria) of Asia. *Palaeontol. Polonica* **33**, 133–181 (1975).
14. Dodson, P. Quantitative aspects of relative growth and sexual dimorphism in *Protoceratops*. *J. Paleontol.* **50**, 929–940 (1976).

15. Coombs, W. P. Juvenile specimens of the ornithischian dinosaur *Psittacosaurus*. *Palaeontology* **25**, 89–107 (1982).
16. Sereno, P. C. in *Dinosaur Systematics, Perspectives and Approaches* (eds Carpenter, K. & Currie, P.) 203–210 (Cambridge Univ. Press, New York, 1990).
17. Chinnery, B. J. & Weishampel, D. B. *Montanoceratops cerorhynchus* (Dinosauria, Ceratopsia) and relationships among basal neoceratopsians. *J. Vert. Paleontol.* **18**, 569–585 (1998).
18. Makovicky, P. J. in *Mesozoic Terrestrial Life* (eds Tanke, D. & Carpenter, K.) 243–262 (Univ. Indiana Press, Bloomington, 2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank Z.-H. Zhou for help during the course of the work, members of the Liaoxi expedition of the IVPP for help in the field, Y.-L. Huo, F.-L. Wang and H.-J. Wang for preparing the specimens, and R.-S. Li for drawings. This work was supported by the Special Funds for Major State Basic Research Projects of China, the National Geographic Society of USA, the Chinese Natural Science Foundation, the Chinese Academy of Sciences and The Field Museum.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.X. (e-mail: xxu@midwest.com.cn).

Remains of *Homo erectus* from Bouri, Middle Awash, Ethiopia

Berhane Asfaw*, W. Henry Gilbert†, Yonas Beyene‡, William K. Hart§, Paul R. Renne¶, Giday WoldeGabriel¶, Elisabeth S. Vrba# & Tim D. White†

* Rift Valley Research Service, P.O. Box 5717, Addis Ababa, Ethiopia

† Department of Integrative Biology and Laboratory for Human Evolutionary Studies, Museum of Vertebrate Zoology, VLSB, University of California, Berkeley, California 94720, USA

‡ Department of Anthropology and Archaeology ARCC, Ministry of Youth, Sports and Culture, Addis Ababa, Ethiopia

§ Department of Geology, Miami University, Oxford, Ohio 45056, USA

¶ Berkeley Geochronology Center, 2455 Ridge Road, Berkeley, California 94709, USA, and Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA

¶ EES-6/MS D462 Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06511, USA

The genesis, evolution and fate of *Homo erectus* have been explored palaeontologically since the taxon's recognition in the late nineteenth century. Current debate¹ is focused on whether early representatives from Kenya and Georgia should be classified as a separate ancestral species (*'H. ergaster'*)^{2–4}, and whether *H. erectus* was an exclusively Asian species lineage that went extinct^{5,6}. Lack of resolution of these issues has obscured the place of *H. erectus* in human evolution. A hominid calvaria and postcranial remains recently recovered from the Dakanihylo Member of the Bouri Formation, Middle Awash, Ethiopia, bear directly on these issues. These ~1.0-million-year (Myr)-old Pleistocene sediments contain abundant early Acheulean stone tools and a diverse vertebrate fauna that indicates a predominantly savannah environment. Here we report that the 'Daka' calvaria's metric and morphological attributes centre it firmly within *H. erectus*. Daka's resemblance to Asian counterparts indicates that the early African and Eurasian fossil hominids represent demes of a widespread palaeospecies. Daka's anatomical intermediacy between earlier and later African fossils provides

evidence of evolutionary change. Its temporal and geographic position indicates that African *H. erectus* was the ancestor of *Homo sapiens*.

The Early Pleistocene Dakanihylo ('Daka') Member comprises 22 to ≤45 m of sediments unconformably atop the Pliocene Hatayae Member of the Bouri Formation⁷ (Fig. 1). These deposits contain abundant archaeological and palaeontological remains embedded in primarily alluvial deposits relating to lakeside beaches or shallow water deposits in distributary channels⁸. Initial interpretations identified Daka sediments as postdating hominid remains and Acheulean artefacts from Bodo⁹. However, the Daka artefacts clearly antedate those at Bodo, and single-crystal ⁴⁰Ar/³⁹Ar dating of a pumiceous unit at the base of the Member gave an age of 1.042 ± 0.009 Myr (ref. 8). The entire Dakanihylo Member is of reverse magnetic polarity, so the minimum age of its palaeoanthropological contents is ~0.8 Myr.

An extensive vertebrate fauna was recovered from the Daka Member. Faunas of this age are rare in Africa. Of 713 identified specimens (see Supplementary Information for faunal list), 377 are bovids, including three new species and two new genera¹⁰. The bovid assemblage is dominated by alcelaphine diversity and abundance not recorded at older African sites. Widespread open grassland habitats are thereby indicated. Adjacent water-margin habitats are evidenced by three *Kobus* species and abundant hippo fossils.

Daka Member archaeological sites are abundant. Bone modifications characteristic of the butchery of large mammals by hominids scar several equid, bovid and hippo postcrania. Lithic assemblages closely conform to African early Acheulean analogues. Handaxes and cleavers are ubiquitous elements, with invasive flake scars and fewer flake removals than later Acheulean counterparts⁸.

A hominid calvaria (BOU-VP-2/66) was discovered *in situ* in Daka Member silty sand by W.H.G. on 27 December 1997. The specimen was orientated base-down, without associated artefacts, encrusted by fossilized root casts. Surface detail is well preserved, with no sign of fluvial transport or surface weathering. Its vault and supraorbitals exhibit perimortem scraping damage; the frontal and parietals bear multiple sets of subparallel striae, each with internal striations. The patterning and morphology of these marks is unusual and inconsistent with cutmarks made by hominids engaged in defleshing activities. We tentatively attribute this damage to animal gnawing.

The calvaria preserves a largely intact base and is only slightly distorted (plastic deformation skews the vault slightly to the individual's left (Fig. 2)). Endocranial capacity is 995 cm³ (measured repeatedly with teff seed). The thick supraorbital tori are strongly arched, with markedly depressed glabellar and supraglabellar regions. Radiographs reveal an asymmetrical frontal sinus extending to the left midorbital level. The frontal squama is bossed at midline and there is weak sagittal keeling there and on the parietals. The mandibular fossa is deep and anteroposteriorly short. Suprameatal and supramastoid crests and angular tori are weak. The damaged mastoids are small. There is no true occipital torus demarcated superiorly by a supratotal sulcus. Rather, the occipital squama rises vertically and curves anteriorly. Viewed posteriorly, the undistorted parietal walls would have been vertical. The cranial vault is smaller and shorter than Olduvai hominid OH-9 and is phenetically similar to the partly described Buia cranium from Eritrea¹¹. Three isolated hominid femora and a proximal tibia were recovered from Daka deposits far removed from the calvaria (Fig. 1). No femur is complete, but all display the marked platymeria and extremely thick midshaft cortex characteristic of *H. erectus*.

The new 'Daka' hominid fossils afford unique insights into unresolved spatial and temporal relationships of *H. erectus*. Most fossils attributed to this taxon came to light in Java, China, Europe and Africa during the twentieth century. Additional genus and species names were proposed before the application of modern

systematics united them under *H. erectus* in the 1960s. The discovery of older Kenyan fossils in the 1970s and the application of cladistic methods in the 1980s produced a variety of phylogenetic and taxonomic assessments of Early and Middle Pleistocene *Homo*. In his 1985 distillation, Delson¹ identified two basic research problems involving *H. erectus*: whether the African fossils were conspecific with the Trinil holotype and other Asian representatives; and whether the species showed stasis or phyletic change through its 1.5-Myr span. Both problems have persisted, and at ~1.0 Myr in the Horn of Africa the new Daka fossils now bear directly on them.

It has been proposed that the name *H. erectus* be restricted to a purported Asian clade (species) exemplified by fossils of disparate antiquity from Trinil, Sangiran, Zhoukoudian and Ngandong^{5,6}. We examined the hypothesis that *H. erectus* was a specifically Asian clade by metrically and cladistically analysing the Daka calvaria. Its cranial metrics (see Supplementary Information) overlap with both Asian and African sample ranges and fail to distinguish the fossil consistently from either sample.

Previous applications of cladistics to *H. erectus* have been criticized on the basis of issues of character independence and variation^{12–14}, as well as the potentially confounding effect of gene flow^{15,16}. We agree with these and other valid cautions^{17,18}. However, to examine the hypothesis that *H. erectus* was a distinct African clade, we experimented with Hennigian parsimony analyses to investigate potential clustering of relevant crania. From recent published literature we compiled the 22 characters most widely and appropriately used in the cladistic analysis of calvarial anatomy in *H. erectus* and close relatives. We divided Early and Middle Pleistocene *Homo* fossils into operational taxonomic units on the basis of the palaeo-demes defined by Howell¹⁹ (these are sets of fossils representing spatially and temporally bounded ‘communities’ below the species level; specimen-by-specimen palaeo-deme content and details of our analysis are presented in Supplementary Information).

Regardless of the software parameters used, or the removal of the later Asian demes, the hypothesis of a deep cladogenesis between African and Asian *H. erectus* is unsupported by our analyses (Fig. 2). Previous cladistic efforts have noted difficulty with the African OH 9 specimen because it consistently aligned with the Asian fossils,

thereby being interpreted as a sort of Tanzanian outpost of ‘Asian’ morphology³ or as an evolutionary intermediate²⁰. The recovery of the Buia and Daka fossils, almost certainly from the same eastern African deme, compounds such problems. Like OH 9, these new African specimens share many derived characters with Asian and European specimens. As a consequence, the cladistic method, regardless of serious questions concerning its applicability here, fails to support the division of *H. erectus* into Asian and African clades. Whether viewed metrically or morphologically, the Daka cranium confirms previous suggestions^{15,16} that geographic subdivision of early *H. erectus* into separate species lineages is biologically misleading, artificially inflating early Pleistocene species diversity. Rather, the Daka calvaria is consistent with the hypothesis of a widespread, moderately polymorphic and polytypic species at ~1.0 Myr (refs 21, 22).

Because it is from a poorly sampled period²³, the newly discovered Daka calvaria is also important for assessing evolutionary mode and tempo in *H. erectus*. A key barrier to such assessment is the lack of chronological control for many Pleistocene Eurasian hominids—a persistent problem that confounds attempts to document patterns of morphological change across time. Chronometric placement is superior in eastern Africa, where the earliest fossils attributed to *H. erectus* (Nariokotome deme) date to ~1.78 Myr (ref. 2). These fossils have smaller braincases than later African species representatives, including Daka. This pattern could also characterize Eurasia, where the earliest hominid fossils from Sangiran, Java²⁴, and Dmanisi, Georgia⁴, might be more primitive than younger regional counterparts (although the chronological placement of most Eurasian fossils is inadequate relative to what is now available for eastern Africa²⁵).

Chronological and anatomical seriation of the African fossils from KNM-ER3733/3883 (Koobi Fora) to OH 9, to Daka/Buia to Bodo²⁶, is now available. In many features, including cranial capacity and its extensive developmental correlates, hominid crania in this eastern African succession comprise a morphocline consistent with the hypothesis that they sample a single evolving lineage. To recognize the basal fossils representing this apparently evolving lineage with the separate species name ‘*H. ergaster*’ is therefore doubtfully necessary or useful². At most, the basal members

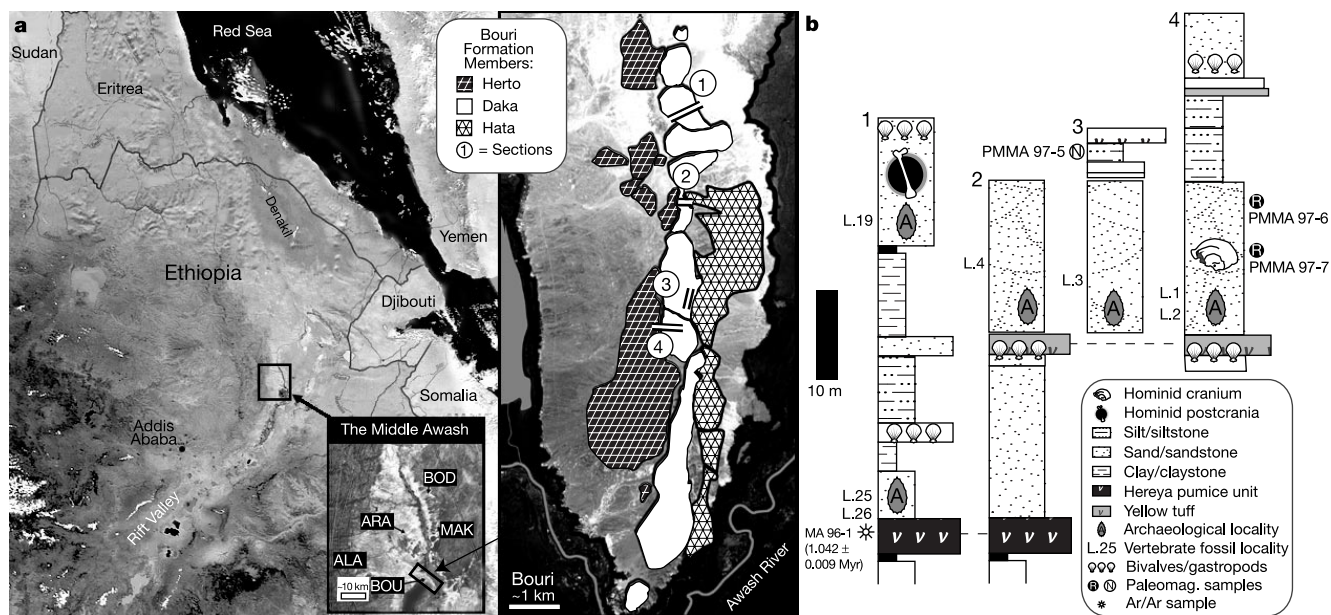


Figure 1 Maps and generalized sections. **a**, Landsat and aerial photographic imagery showing the location of the Middle Awash study area, its major hominid-bearing localities, and the exposure of Bouri Formation sediments. The 1.0-Myr Daka Member of this formation is shown in white. **b**, Measured stratigraphic sections of this Member show the

placement of Acheulean artefacts, hominid fossils, radiometric dates and geomagnetic polarity determinations (normal polarity in section 3 and cap of section 4 overlie the Daka Member).

of the *H. erectus* lineage should be recognized taxonomically as a chrono-subspecies (*H. erectus ergaster*). Suggestions that *H. ergaster* itself contains multiple species, even in a single locality²⁷, seem completely unsupported by the data⁴.

The origins of the widespread, polymorphic, Early Pleistocene *H. erectus* lineage remain elusive. The marked contrasts between any potential ancestor (*Homo habilis* or other) and the earliest known *H. erectus* might signal an abrupt evolutionary emergence some time before its first known appearance in Africa at ~1.78 Myr. Uncertainties surrounding the taxon's appearance in Eurasia and southeast Asia make it impossible to establish accurately the time or place of origin for *H. erectus*. Available evidence is insufficient to detect the direction of its geographic dispersal. Given new perspectives afforded by the discoveries at Dmanisi in Eurasia, the assumption that the earliest *H. erectus* populations emigrated from Africa to Eurasia^{14,23,28,29}, rather than invading Africa from Eurasia³⁰, is premature. Whatever its time and place of origin, and direction of spread, this species dispersed widely, and possibly abruptly, before 1.5 Myr. The Daka calvaria indicates that by 1 Myr the taxon had colonized much of the Old World without speciating—a finding of considerable biogeographic and behavioural significance. Through time and across its eastern hemispheric range, the technologies employed by this taxon ranged from Oldowan to Acheulean.

The resemblance of the Daka calvaria to Asian representatives of *H. erectus*, and its morphological intermediacy between earlier and later African specimens, provide strong evidence that it samples a widely distributed lineage that evolved during the million years after its Pliocene origin. The phylogenetic unity of *H. erectus* did not persist indefinitely. By ~0.5 Myr, very different hominid crania in Africa and Asia (Bodo versus Zhoukoudian) indicate that a hominid speciation event might have occurred in circum-Daka times²⁶. Further sampling of the Daka to Bodo transition in Africa is needed to examine the rate of morphological change, as well as the hypothesis that the fractionation of *H. erectus* might have been related to the ~0.95-Myr onset of large magnitude global climatic oscillations. □

Received 6 December 2001; accepted 8 February 2002.

1. Delson, E. Palaeobiology and age of African *Homo erectus*. *Nature* **316**, 762–763 (1985).
2. Walker, A. C. & Leakey, R. E. F. *The Nariokotome Homo erectus Skeleton* (Harvard Univ. Press, Cambridge, Massachusetts, 1993).
3. Wood, B. A. *Koobi Fora Research Project Vol. 4* (Clarendon Press, Oxford, 1991).
4. Gabunia, L., Vekua, A. & Lordkipanidze, D. New human fossils from Dmanisi, eastern Georgia. *Archaeol. Ethnol. Anthropol. Eurasia* **2**(6), 128–139 (2001).
5. Andrews, P. J. An alternative interpretation of the characters used to define *Homo erectus*. *Courier Forschungsinstitut Senckenberg* **69**, 167–175 (1984).
6. Groves, C. P. *A Theory of Human and Primate Evolution* (Clarendon Press, Oxford, 1989).
7. DeHeinzelin, J. et al. Environment and behavior of 2.5-million-year-old Bouri hominids. *Science* **284**, 625–629 (1999).
8. DeHeinzelin, J., Clark, J. D., Schick, K. D. & Gilbert, W. H. (eds) *The Acheulean and the Plio-Pleistocene Deposits of the Middle Awash Valley, Ethiopia* (Royal Museum of Central Africa, Tervuren, Belgium; *Ann. Sci. Geol.*, **104** (2000)).
9. Kalb, J. et al. Geology and stratigraphy of Neogene deposits, Middle Awash Valley, Ethiopia. *Nature* **298**, 17–25 (1982).
10. Vrba, E. S. New fossils of Alcelaphini and Caprinae (Bovidae: Mammalia) from Awash, Ethiopia, and phylogenetic analysis of Alcelaphini. *Palaeont. Afr.* **34**, 127–198 (1997).
11. Abbate, E. et al. A one-million-year-old *Homo* cranium from the Danakil (Afar) depression of Eritrea. *Nature* **393**, 458–460 (1998).
12. Brauer, G. & Mbua, E. *Homo erectus* features used in cladistics and their variability in Asian and African hominids. *J. Hum. Evol.* **22**, 79–108 (1992).
13. Kennedy, G. E. On the autapomorphic traits of *Homo erectus*. *J. Hum. Evol.* **20**, 375–412 (1991).
14. Kramer, A. Human taxonomic diversity in the Pleistocene: Does *Homo erectus* represent multiple hominid species? *Am. J. Phys. Anthropol.* **91**, 161–171 (1993).
15. Harrison, T. in *Species, Species Concepts, and Primate Evolution* (eds Kimbel, W. H. & Martin, L. B.) 345–371 (Plenum, New York, 1993).
16. Turner, A. & Chamberlain, A. Speciation, morphological change and the status of African *Homo erectus*. *J. Hum. Evol.* **18**, 115–130 (1989).
17. Lovejoy, C. O., Cohn, M. J. & White, T. D. Morphological analysis of the mammalian postcranium: A developmental perspective. *Proc. Natl Acad. Sci. USA* **96**, 13247–13252 (1999).
18. Trinkaus, E. Cladistics and the hominid fossil record. *Am. J. Phys. Anthropol.* **83**, 1–11 (1990).
19. Howell, F. C. Paleo-demes, species clades, and extinctions in the Pleistocene hominid record. *J. Anthropol. Res.* **55**, 191–243 (1999).
20. Martinez, I. & Arsuaga, J. L. The temporal bones from Sima de los Huesos Middle Pleistocene site (Sierra de Atapuerca, Spain). A phylogenetic approach. *J. Hum. Evol.* **33**, 283–318 (1997).
21. Rightmire, G. P. *The Evolution of Homo erectus: Comparative Anatomical Studies of an Extinct Human Species* (Cambridge Univ. Press, Cambridge, 1990).
22. Rightmire, G. P. Evidence from facial morphology for similarity of Asian and African representatives of *Homo erectus*. *Am. J. Phys. Anthropol.* **106**, 61–85 (1998).
23. Aguirre, E. Poor fossil record and major changes around 1 MaBP. *Hum. Evol.* **15**, 51–62 (2000).
24. Larick, R. et al. Early Pleistocene ⁴⁰Ar/³⁹Ar ages for Bapang Formation hominins, Central Java, Indonesia. *Proc. Natl Acad. Sci. USA* **98**, 4866–4871 (2001).
25. Olsen, J. W. New light on the earliest occupation of east Asia. *Anthropologie* **37**, 89–96 (1999).
26. Rightmire, G. P. The human cranium from Bodo, Ethiopia: Evidence for speciation in the Middle Pleistocene. *J. Hum. Evol.* **31**, 21–39 (1996).
27. Schwartz, J. H. Taxonomy of the Dmanisi crania. *Science* **289**, 55 (2000).
28. Ambrose, S. H. Paleolithic technology and human evolution. *Science* **291**, 1748–1753 (2001).
29. Bar-Yosef, O. Lower Paleolithic sites in South-western Asia: Evidence for 'Out of Africa' movements. *Anthropologie* **37**, 51–69 (1999).
30. Clarke, R. Out of Africa and back again. *Int. J. Anthropol.* **15**, 185–189 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

This paper is dedicated to the late J. Desmond Clark who initiated this research at Bouri in 1981. We thank A. Almquist, A. Asfaw, M. Asnake, T. Asseberew, D. Brill, D. DeGusta, J. DeHeinzelin, A. Getty, Y. Haile-Selassie, B. Latimer, C. Pehlevan, K. Schick, S. Simpson, P. Snow, G. Suwa and Y. Zeleke for fieldwork and analytical studies. Electron microprobe and other support from the Earth Environmental Sciences Division, Los Alamos National Laboratory. Thanks to G. Suwa, D. DeGusta, C. Feibel, F. C. Howell, C. O. Lovejoy, F. Bibi, T. Stidham, J. Parham, J.-R. Boissarie and H. Saegusa for thoughtful review and/or assistance. We thank the Ministry of Youth, Sports and Culture, the Authority for Research and Conservation of the Cultural Heritage, and the National Museum of Ethiopia for

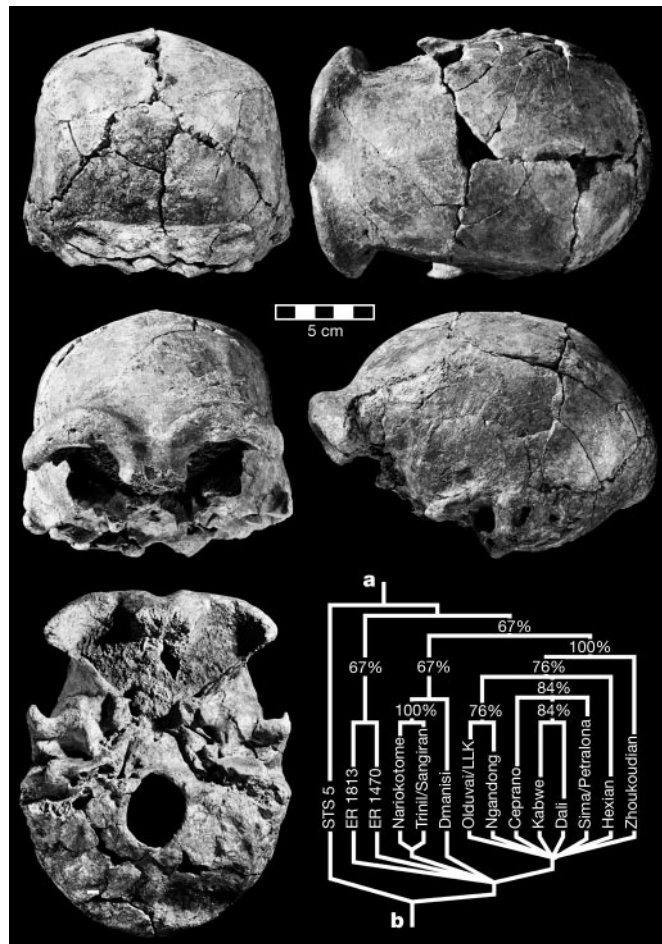


Figure 2 Views of the Daka calvaria and cladograms representing majority rule (a) and strict consensus (b) of the 75 most parsimonious phylogenies generated by PAUP 4.0b8 (Sinauer). The Daka cranium is in the 'Olduvai/LLK' deme. See the text and Supplementary Information for analytical details. Cladogram b is more conservative, showing the nodes that are stable across all equally parsimonious phylogenies. These branching diagrams are presented only to illustrate the limitations of a cladistic approach to phylogeny among this set of fossils, and the lack of calvarial evidence for a deep phylogenetic division between the African and Asian fossils.

permissions, and the Afar Regional Government and the Afar people of the Middle Awash, particularly H. Elema. Many additional individuals contributed. This research was supported by the NSF (USA) and the Institute of Geophysics and Planetary Physics (University of California at Los Alamos National Laboratory), with additional contributions by the Graduate School and Hampton Fund for International Initiatives, Miami University.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.A. (e-mail: ramid@telecom.net.et).

Genetic cost of reproductive assurance in a self-fertilizing plant

Christopher R. Herlihy & Christopher G. Eckert

Department of Biology, Queen's University, Kingston, Ontario, Canada K7L 3N6

The transition from outcrossing to self-fertilization is one of the most common evolutionary trends in plants¹. Reproductive assurance, where self-fertilization ensures seed production when pollinators and/or potential mates are scarce, is the most long-standing and most widely accepted explanation for the evolution of selfing^{2–8}, but there have been few experimental tests of this hypothesis. Moreover, many apparently adaptive floral mechanisms that ensure the autonomous production of selfed seed might use ovules that would have otherwise been outcrossed. This seed discounting is costly if selfed offspring are less viable than their outcrossed counterparts, as often happens. The fertility benefit of reproductive assurance has never been examined in the light of seed discounting. Here we combine experimental measures of reproductive assurance with marker-gene estimates of self-fertilization, seed discounting and inbreeding depression to show that, during 2 years in 10 Ontario populations of *Aquilegia canadensis* (Ranunculaceae), reproductive assurance through self-fertilization increases seed production, but this benefit is greatly outweighed by severe seed discounting and inbreeding depression.

Most theoretical models for the evolution of self-fertilization have focused on two opposing consequences of selfing. Self-fertilization is beneficial because an outcrosser only sires the offspring of other individuals, whereas a selfer also sires its own offspring. An allele that causes selfing can therefore increase its own transmission

by up to 50% in an outcrossing population⁹. In contrast, self-fertilization is detrimental when selfed offspring suffer reduced viability because of inbreeding depression^{2,10}. These models typically predict that mating systems should evolve towards two evolutionarily stable endpoints: predominant selfing should evolve when the fitness of selfed progeny is at least half that of outcrossed progeny, whereas predominant outcrossing should evolve when selfed progeny are less than half as fit as outcrossed progeny^{7,8,10}. However, at least one-third of plant species for which the mating system has been quantified practise a broad mixture of selfing and outcrossing¹¹. Moreover, many species that practise mixed mating exhibit inbreeding depression that is strong enough to outweigh the gene-transmission advantage of selfing^{12,13}. This suggests that other selective factors also influence the evolution of self-fertilization.

The reproductive assurance hypothesis resolves this disparity between theoretical predictions and empirical results by explaining how self-fertilization can be advantageous in species with strong inbreeding depression. Self-fertilization is additionally beneficial if it enables the production of seeds when pollinators and/or potential mates are scarce⁴. If opportunities for outcrossing are limited, selfing can be selected even if inbreeding depression is strong^{14–16}. Although the reproductive assurance hypothesis, first championed by Darwin², is now widely accepted as an explanation for the evolution of selfing, it has never been subject to a rigorous experimental test. A comprehensive test of this hypothesis must satisfy three requirements. First, selfing need not provide reproductive assurance every year in every population for it to be favoured¹⁶, so it is necessary to examine the fitness consequences of selfing in multiple populations, or across multiple years within individual populations. Most studies of reproductive assurance, including previous work on *Aquilegia canadensis*, involved single populations in single reproductive seasons^{17,18} (but see refs 19, 20). Second, self-fertilization can result from both within-flower (autogamous) and between-flower (geitonogamous) self-pollination, yet only autogamous selfing is likely to provide reproductive assurance¹⁴. The reproductive assurance hypothesis must be tested by estimating the specific fitness consequences of autogamous selfing^{14,15}. Third, selfing can increase seed production when the opportunity for outcrossing is limited, but the benefits of reproductive assurance will erode if ovules that would have been outcrossed are self-fertilized instead, and selfed offspring are inferior owing to inbreeding depression¹⁴. The absence of seed discounting is an integral component of the reproductive assurance hypothesis, yet the few experimental tests performed so far have determined only whether selfing increases total seed production and not whether a gain in selfed seeds trades off with a loss of outcrossed seeds¹⁵. The benefit of reproductive assurance cannot be determined without accounting for fitness costs associated with seed discounting and inbreeding

Table 1 Seed production and proportion of seeds produced through self-fertilization in intact and emasculated flowers

Year	Population	Seeds per carpel (mean $\pm$ s.e.m.)		Level of selfing (mean $\pm$ s.e.m.)		<i>F</i> (mean $\pm$ s.e.m.)	δ (mean $\pm$ s.e.m.)
		Intact	Emasculated	Intact	Emasculated		
1998	QCA2	11.40 $\pm$ 1.70	9.32 $\pm$ 1.92	1.000 $\pm$ 0.147	0.784 $\pm$ 0.144	–	–
	QLL1	10.35 $\pm$ 1.85	9.66 $\pm$ 1.27	0.810 $\pm$ 0.075	0.598 $\pm$ 0.113	–	–
	QSP1	8.71 $\pm$ 1.08	5.86 $\pm$ 1.37	0.839 $\pm$ 0.053	0.526 $\pm$ 0.141	–	–
1999	QCA2	14.06 $\pm$ 1.15	13.90 $\pm$ 1.15	0.765 $\pm$ 0.081	0.348 $\pm$ 0.113	–0.140 $\pm$ 0.117	1.075 $\pm$ 0.061
	QFP1	15.89 $\pm$ 1.34	13.22 $\pm$ 1.39	0.695 $\pm$ 0.068	0.478 $\pm$ 0.107	0.076 $\pm$ 0.092	0.928 $\pm$ 0.120
	QLL1	14.72 $\pm$ 1.00	14.37 $\pm$ 0.98	0.751 $\pm$ 0.057	0.250 $\pm$ 0.101	0.036 $\pm$ 0.105	0.975 $\pm$ 0.083
	QLL2	15.46 $\pm$ 1.16	14.00 $\pm$ 1.18	0.628 $\pm$ 0.070	0.365 $\pm$ 0.098	–0.070 $\pm$ 0.079	1.078 $\pm$ 0.111
	QMT1	14.37 $\pm$ 1.13	11.50 $\pm$ 1.42	0.731 $\pm$ 0.076	0.447 $\pm$ 0.106	–	–
	QOR1	14.72 $\pm$ 1.01	12.87 $\pm$ 1.10	0.670 $\pm$ 0.056	0.206 $\pm$ 0.081	0.122 $\pm$ 0.064	0.864 $\pm$ 0.087
	QPG1	14.96 $\pm$ 0.99	12.91 $\pm$ 1.30	0.827 $\pm$ 0.056	0.792 $\pm$ 0.059	0.060 $\pm$ 0.099	0.973 $\pm$ 0.051
	QRR1	13.06 $\pm$ 1.52	10.15 $\pm$ 1.70	0.911 $\pm$ 0.061	0.538 $\pm$ 0.102	–	–
	QSP1	13.84 $\pm$ 1.17	11.48 $\pm$ 1.25	0.740 $\pm$ 0.077	0.642 $\pm$ 0.101	–0.074 $\pm$ 0.079	1.048 $\pm$ 0.052
	QTF1	13.18 $\pm$ 1.17	12.41 $\pm$ 0.92	0.628 $\pm$ 0.048	0.178 $\pm$ 0.084	0.060 $\pm$ 0.094	0.925 $\pm$ 0.097

Means for seed production are based on 6–43 plants (mean 24) per treatment per population. Emasculatioin reduced seed production in all populations (range 1.2–32.7% reduction; mean $\pm$ s.e.m. 13.8 $\pm$ 2.4%; paired *t*-test, 1-tailed *P* < 0.0001). Estimates of selfing, the parental inbreeding coefficient (*F*) and inbreeding depression (δ) are based on 9–57 progeny arrays (mean 30) per treatment per population. Emasculatioin also reduced the level of self-fertilization in all populations (range 4.2–71.6% reduction; mean $\pm$ s.e.m. 39.8 $\pm$ 5.9%; paired *t*-test, 1-tailed *P* < 0.01).

depression. No previous test of the reproductive assurance hypothesis has met all three requirements.

We tested the reproductive assurance hypothesis during two flowering seasons in 10 natural populations of *A. canadensis* near the Queen's University Biological Station, Chaffey's Lock, Ontario, Canada. Populations of *A. canadensis* from throughout the species' geographical range exhibit high levels (~75%) of self-fertilization (ref. 21 and C.R.H., B. Ozimec and C.G.E., unpublished observations), and several aspects of the species' ecology indicate reproductive assurance as an adaptive explanation for selfing. First, *A. canadensis* typically occurs in small patchy populations where the availability of outcross mates might be limited^{21,22}. Second, it flowers in the spring when pollinator service (usually bees and hummingbirds) might be unpredictable owing to inclement weather²³. Third, exclusion experiments showed that *A. canadensis* produces a full complement of seeds by autonomous selfing in the absence of pollinators^{17,21,23}. Last, inbreeding depression in *A. canadensis* is more than strong enough to outweigh the gene-transmission advantage of selfing²¹.

To determine the extent to which autogamous self-fertilization results in reproductive assurance versus seed discounting, we emasculated individual flowers, thereby eliminating the capacity for autogamy. We estimated reproductive assurance (R) as the difference in seed production between intact and emasculated flowers, and combined these seed production data with estimates of self-fertilization to quantify the level of autogamous selfing (a) and seed discounting (D). Inbreeding depression (δ) was estimated by comparing the change in inbreeding coefficient between seeds and mature plants. If reproductive assurance is an adaptive explanation for high levels of selfing in *A. canadensis*, intact flowers should produce more seed than emasculated flowers. We also expect that the level of autogamous selfing should be positively correlated with the increase in seed production afforded by selfing among populations. Individuals in populations in which pollinators and/or mates are scarcest will exhibit the largest increase in seed production via reproductive assurance and, as a result, the highest levels of autogamous self-fertilization. Finally, the adaptive significance of reproductive assurance depends strongly on the extent to which selfing pre-empts seeds that could have otherwise been outcrossed. The ultimate fitness consequences of the trade-off between selfed and outcrossed seeds depends on the strength of inbreeding depression¹⁴.

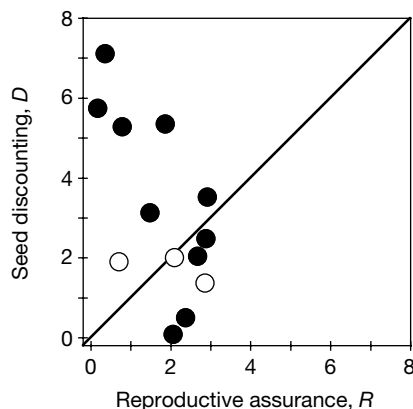


Figure 1 Trade-off between the gain in seed production through autogamous selfing (reproductive assurance, R) and the loss of outcrossed seeds (seed discounting, D). Open and filled points are population means from 1998 and 1999, respectively. Points lying above the diagonal are populations in which the gain in total seed production was more than offset by the loss of outcrossed seed. Mean ($\pm$ s.e.m.) reproductive assurance was 1.77 ± 0.27 seeds per carpel, whereas the loss of outcrossed seeds was 3.13 ± 0.60 seeds per carpel.

Experimentally preventing autogamous self-fertilization reduced seed production in all populations of *A. canadensis* (mean reduction 14%), indicating that autogamy provides substantial reproductive assurance (Table 1). However, emasculation reduced self-fertilization to a much larger extent (mean reduction 40%; Table 1) indicating that autogamy also results in seed discounting. The average gain in seed production ($R = +1.77 \pm 0.27$ seeds per carpel) was associated with a much greater loss of outcrossed seed ($D = -3.13 \pm 0.60$ seeds per carpel; Fig. 1). Contrary to expectations, reproductive assurance was not positively correlated with autogamous selfing among populations (Fig. 2a). Instead, higher levels of autogamous selfing simply resulted in greater losses of outcrossed seed (Fig. 2b). All estimates of inbreeding depression were high ($\delta = 1 - \text{relative fitness of selfed progeny}$; mean $\pm$ s.e.m. = 0.98 ± 0.03 ; Table 1) and very similar to those from 10 additional Ontario populations²¹ and several populations from throughout the species' range (C.R.H. and C.G.E., unpublished observations).

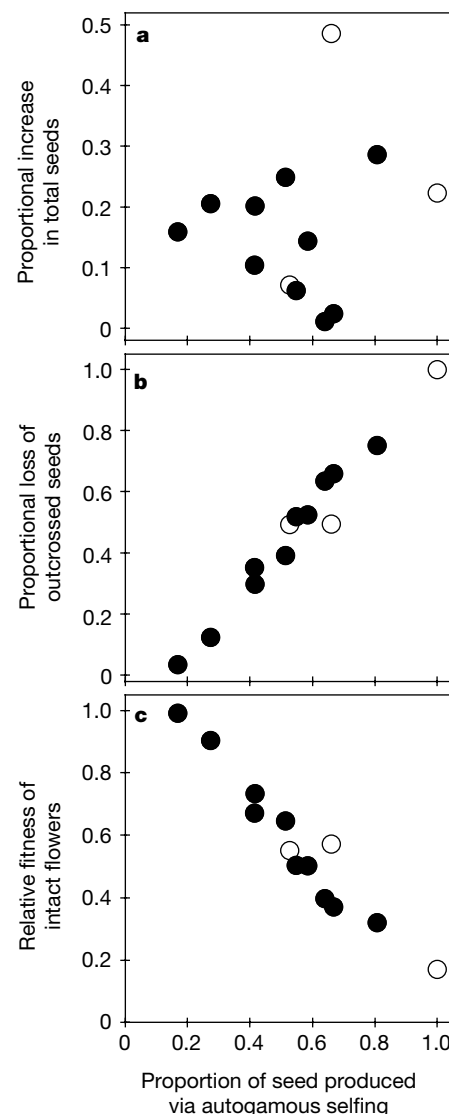


Figure 2 Cost-benefit analysis of variation in autogamous selfing among populations. **a**, Reproductive assurance (R) is not correlated with autogamous selfing among populations ($r = +0.16$, $P = 0.59$). **b**, The proportional loss of outcrossed seed production increases strongly with autogamous selfing among populations ($r = +0.98$, $P < 0.0001$). **c**, The relative fitness of intact flowers decreases strongly with selfing among populations ($r = -0.96$, $P < 0.0001$).

Because selfed progeny rarely survive to reproduce, seed discounting is very costly. We evaluated the overall fitness consequence of autogamy by combining the benefits of reproductive assurance and gene transmission with the costs of seed discounting and inbreeding depression to estimate the genetic contribution of intact versus emasculated flowers through seed to the next generation. The relative fitness of intact flowers was less than 1 in all populations (range 0.17–0.99; mean $\pm$ s.e.m. = 0.56 ± 0.06) and decreased strongly with increased autogamous selfing (Fig. 2c). The fertility advantage of autogamous selfing is more than outweighed by substantial seed discounting combined with strong inbreeding depression. Self-fertilization seems strongly disadvantageous in these populations of *A. canadensis*.

Modifications to floral morphology or development that reduce seed discounting but allow reproductive assurance should be favoured by selection in *A. canadensis*. Traits that promote outcrossing before ovules are selfed would reduce seed discounting^{14,15}. There are many descriptive studies of potential delayed selfing mechanisms²⁴ but no investigation of whether they actually uncouple seed discounting from reproductive assurance. Protogyny (presenting receptive stigmas to receive outcross pollen before self pollen is shed from anthers), is a very common floral trait that is widely viewed as a delayed selfing mechanism. Flowers of several species of *Aquilegia*, including *A. canadensis*, have been described as protogynous, but experiments revealed that, in *A. canadensis*, stigmas become receptive only when anthers release self pollen²⁵. Self-pollination and cross-pollination probably occur simultaneously. Moreover, field experiments in which pollen presentation was delayed until well after the onset of stigma receptivity failed to show that protogyny would reduce seed discounting²⁵. The most likely mechanism for avoiding seed discounting in this species might not function effectively in natural populations.

Trade-offs are thought to have a strong influence on the evolution of many aspects of life history. In theory, gamete discounting is a potentially important trade-off in the evolution of selfing, and a weak or non-existent trade-off between self and outcross seed production is a key feature of the reproductive assurance hypothesis^{14,15}. Our results challenge this widely held explanation for the evolution of selfing. *A. canadensis* occurs in circumstances that should strongly favour reproductive assurance. Indeed, autonomous selfing substantially increases total seed production. However, this is more than outweighed by a loss of high-quality outcross seed. This is the first empirical demonstration of seed discounting as a potent selective force in natural populations. Seed discounting might be especially costly in sequentially flowering perennials such as *A. canadensis* because it involves the pre-emption of resources as well as ovules that could have been invested in outcrossed seed. Recent theory suggests that the within-flower seed discount shown here might be further manifested in the reduced production of outcrossed seed by future flowers, both within and between reproductive seasons²⁶. This is supported by experimental data from *A. canadensis* showing that resources not invested in seed production by current flowers are reallocated to increase the fecundity of later flowers (A. Kliber and C.G.E., unpublished observations). The potential reverberations of the trade-off between selfed and outcrossed seed production through the life cycle emphasize the importance of considering the evolution of mating systems in the broader context of resource allocation and life history²⁷. □

Methods

Experimental emasculations were performed in 9 populations in 1998, and 12 in 1999. Removing anthers does not have any confounding effects in terms of either damage to flowers or shortened floral lifespan^{17,25}. Emasculations do not greatly reduce the attractiveness of flowers to pollinators, although some decrease in outcross pollination as a by-product of emasculations does not alter the interpretation of our results. To minimize confounding variation in seed production and selfing among flowers owing to plant size and flower-position effects, we used pairs of plants with the same numbers of flowers, and focal flowers in the same position in the flowering sequence. In each population we used 40

pairs of plants in 1998, and 50–100 in 1999. Droughts reduced sample sizes below that required to estimate self-fertilization for 6 populations in 1998 and 2 in 1999. We include data only from populations for which we reliably estimated both seed production and selfing. For each population, estimates of mating parameters were based on progeny arrays of 10–30 seeds collected from each of 9–57 (mean 30) plants per treatment and assayed for two polymorphic allozyme loci²¹. The proportion of seeds produced through selfing (*s*) and the parental inbreeding coefficient (*F*) were estimated for the population as a whole and for both intact and emasculated flowers separately with the use of maximum likelihood²⁸, and s.e.m. values were derived by bootstrapping (as described in ref. 21).

The gain in seed production via autogamy (reproductive assurance, *R*) was calculated for each population as the difference between the mean seed production of intact flowers (*P_i*) and that of emasculated flowers (*P_e*)^{14,29}. The loss of outcrossed seed (*D*) was calculated as $P_e(1 - s_e) - P_i(1 - s_i)$, where *s_i* and *s_e* are estimates of selfing in intact and emasculated flowers, respectively. Autogamous selfing (*a*) was calculated (as in ref. 29) as

$$a = \frac{s_i - s_e}{1 - s_i}$$

Inbreeding depression ($\delta = 1 - \text{relative fitness of selfed progeny}$) was estimated from the change in inbreeding coefficient between seeds and adults as

$$\delta = 1 - \left(\frac{2(1 - s)F}{s(1 - F)} \right),$$

where *s* and *F* were estimated from intact control plants^{21,30}. Several populations that produced unreliable estimates of δ because of a low number of progeny arrays (<25) were excluded from this analysis. This method of estimating δ assumes that *s* and *F* do not fluctuate over time within individual populations or for the species as a whole. However, the only potential violation of these assumptions by populations of *A. canadensis* will underestimate δ (refs 21, 30) and therefore does not alter the interpretation of our results.

The relative fitnesses of intact and emasculated flowers were estimated as W_i/W_e , where *W* is the relative genetic contribution to reproductive plants in the next generation, calculated as the sum of the number of seeds (*P*) produced through selfing (*s*) and outcrossing ($1 - s$) weighted by their relative viability (1 for outcrossed and $1 - \delta$ for selfed) and the genetic representation of the maternal parent (1 for outcrossed and 2 for selfed). Thus, $W_e = (1 - s_e)P_e + s_e P_e 2(1 - \delta)$ and $W_i = (1 - s_i)P_i + s_i P_i 2(1 - \delta)$. The value of δ used in the calculations is the mean estimate from eight populations studied in 1999 (Table 1).

Received 11 September; accepted 21 December 2001.

1. Stebbins, G. L. *Flowering Plants: Evolution above the Species Level* (Belknap, Cambridge, Massachusetts, 1974).
2. Darwin, C. R. *The Effects of Cross and Self-fertilization in the Vegetable Kingdom* (John Murray, London, 1876).
3. Baker, H. G. Self-compatibility and establishment after 'long-distance' dispersal. *Evolution* **9**, 347–348 (1955).
4. Jain, S. K. The evolution of inbreeding in plants. *Annu. Rev. Ecol. Syst.* **7**, 69–95 (1976).
5. Lloyd, D. G. in *Demography and Evolution in Plant Populations* (ed. Solbrig, O. T.) 67–88 (Blackwell, Oxford, 1980).
6. Yahara, T. Graphical analysis of mating system evolution in plants. *Evolution* **46**, 557–561 (1992).
7. Jarne, P. & Charlesworth, D. The evolution of the selfing rate in functionally hermaphroditic plants and animals. *Annu. Rev. Ecol. Syst.* **24**, 441–466 (1993).
8. Holsinger, K. E. Pollination biology and the evolution of mating systems in flowering plants. *Evol. Biol.* **29**, 107–149 (1996).
9. Fisher, R. A. Average excess and average effect of a gene substitution. *Ann. Eugen.* **11**, 53–63 (1941).
10. Charlesworth, D. & Charlesworth, B. Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Syst.* **18**, 237–268 (1987).
11. Barrett, S. C. H., Harder, L. D. & Worley, A. C. The comparative biology of pollination and mating in flowering plants. *Phil. Trans. R. Soc. Lond. B* **351**, 1271–1280 (1996).
12. Lande, R., Schemske, D. W. & Schultz, S. T. High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. *Evolution* **48**, 965–978 (1994).
13. Husband, B. C. & Schemske, D. W. Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* **50**, 54–70 (1996).
14. Lloyd, D. G. Self- and cross-fertilization in plants. II. The selection of self-fertilization. *Int. J. Plant Sci.* **153**, 370–380 (1992).
15. Schoen, D. J., Morgan, M. T. & Bataillon, T. How does self-pollination evolve? Inferences from floral ecology and molecular genetic variation. *Phil. Trans. R. Soc. Lond. B* **351**, 1281–1290 (1996).
16. Schoen, D. J. & Brown, A. H. D. Whole- and part-flower self-pollination in *Glycine clandestina* and *G. argyrea* and the evolution of autogamy. *Evolution* **45**, 1665–1674 (1991).
17. Eckert, C. G. & Schaefer, A. Does self-pollination provide reproductive assurance in wild columbine, *Aquilegia canadensis* (Ranunculaceae)? *Am. J. Bot.* **85**, 919–924 (1998).
18. Cruden, R. W. & Lyon, D. L. in *The Evolutionary Ecology of Plants* (eds Bock, J. H. & Linhart, Y. B.) 171–208 (Westview Press, Boulder, Colorado, 1989).
19. Piper, J. G., Charlesworth, B. & Charlesworth, D. A high rate of self-fertilization and increased seed fertility of homostyle primroses. *Nature* **310**, 50–51 (1984).
20. Piper, J. G., Charlesworth, B. & Charlesworth, D. Breeding system evolution in *Primula vulgaris* and the role of reproductive assurance. *Heredity* **56**, 207–217 (1986).
21. Routley, M. B., Mavraganis, K. & Eckert, C. G. Effect of population size on the mating system in a self-compatible, autogamous plant, *Aquilegia canadensis* (Ranunculaceae). *Heredity* **82**, 518–528 (1999).
22. Mavraganis, K. & Eckert, C. G. Effect of population size and isolation on reproductive output of *Aquilegia canadensis* (Ranunculaceae). *Oikos* **95**, 533–546 (2001).

23. Macior, L. W. The pollination of vernal angiosperms. *Oikos* **30**, 452–460 (1978).
24. Lloyd, D. G. & Schoen, D. J. Self- and cross-fertilization in plants. I. Functional dimensions. *Int. J. Plant Sci.* **153**, 358–369 (1992).
25. Griffin, S. R., Mavraganis, K. & Eckert, C. G. Experimental analysis of protogyny in *Aquilegia canadensis*. *Am. J. Bot.* **87**, 1246–1256 (2000).
26. Morgan, M. T., Schoen, D. J. & Bataillon, T. M. The evolution of self-fertilization in perennials. *Am. Nat.* **150**, 618–638 (1997).
27. Morgan, M. T., and Schoen, D. J. The role of theory in an emerging new plant reproductive biology. *Trends Ecol. Evol.* **12**, 231–234 (1997).
28. Ritland, K. A series of FORTRAN computer programs for estimating plant mating systems. *J. Hered.* **81**, 235–237 (1990).
29. Schoen, D. J. & Lloyd, D. G. Self- and cross-fertilization in plants. III. Methods for studying modes and functional aspects of self-fertilization. *Int. J. Plant Sci.* **153**, 381–393 (1992).
30. Ritland, K. Inferences about inbreeding depression based on changes of the inbreeding coefficient. *Evolution* **44**, 1230–1241 (1990).

Acknowledgements

We thank M. Bhardwaj, S. Griffin and A. Kliber for help with the field work; C. Muis and M. Bhardwaj for help in the laboratory; and S. Barrett, K. Holsinger, A. Kliber, B. Montgomerie and M. Morgan for comments on the manuscript. This work was supported in the field by the Queen's University Biological Station and by a research grant from the Natural Sciences and Engineering Research Council of Canada to C.G.E.

Correspondence and requests for materials should be addressed to C.G.E.
(e-mail: eckertc@biology.queensu.ca).

Reduced adaptation of a non-recombining neo-Y chromosome

Doris Bachtrog & Brian Charlesworth

Institute of Cell, Animal and Population Biology, University of Edinburgh,
West Mains Road, Edinburgh EH9 3JT, UK

Sex chromosomes are generally believed to have descended from a pair of homologous autosomes. Suppression of recombination between the ancestral sex chromosomes led to the genetic degeneration of the Y chromosome¹. In response, the X chromosome may become dosage-compensated^{1,2}. Most proposed mechanisms for the degeneration of Y chromosomes involve the rapid fixation of deleterious mutations on the Y¹. Alternatively, Y-chromosome degeneration might be a response to a slower rate of adaptive evolution, caused by its lack of recombination³. Here we report patterns of DNA polymorphism and divergence at four genes located on the neo-sex chromosomes of *Drosophila miranda*. We show that a higher rate of protein sequence evolution of the neo-X-linked copy of *Cyclin B* relative to the neo-Y copy is driven by positive selection, which is consistent with the adaptive hypothesis for the evolution of the Y chromosome³. In contrast, the neo-Y-linked copies of *even-skipped* and *roundabout* show an elevated rate of protein evolution relative to their neo-X homologues, probably reflecting the reduced effectiveness of selection against deleterious mutations in a non-recombining genome¹. Our results provide evidence for the importance of sexual recombination for increasing and maintaining the level of adaptation of a population.

Well-studied sex chromosome systems, such as those of humans or *Drosophila melanogaster*, are very ancient and show few signs of their evolutionary origins^{4,5}. In *D. miranda*, a neo-sex chromosome system has resulted from a recent chromosome fusion between the Y chromosome and an autosome (Muller's element C (ref. 6); Fig. 1). The fused autosome, the neo-Y chromosome, is transmitted patrilineally. Its homologue segregates with the X chromosome, forming the neo-X chromosome. The closest relatives of *D. miranda*, *D. pseudoobscura* and *D. persimilis* (from which it diverged about 2 million years ago⁷), lack the fusion, setting an upper limit to the age of the neo-sex chromosomes. Because male *Drosophila* have

achiasmatic meiosis⁸, the neo-Y never recombines, and so is subject to the same evolutionary forces responsible for the degeneration of 'true' Y chromosomes. Indeed, the neo-sex chromosomes of *D. miranda* show many characteristics of the much older true sex chromosomes. The neo-Y has degenerated substantially⁹, whereas the neo-X chromosome has evolved partial dosage compensation². There is compelling evidence that much of the human Y chromosome is derived from a single autosomal region added to the original Y chromosome¹⁰; that is, it is also a degenerate neo-Y chromosome.

The different evolutionary processes that might be involved in Y-chromosome degeneration leave different footprints of variation and evolution at the molecular level¹, and so can be examined empirically. Here we report levels of polymorphism and divergence at four genes, *Cyclin B* (*CycB*), *roundabout* (*robo*), *engrailed* (*eng*) and *even-skipped* (*eve*), located on both the neo-X and neo-Y chromosomes of *D. miranda*. The use of degenerate primers for PCR, and screening of a genomic library, allowed the isolation of these genes from *D. miranda* (see Methods). *In situ* hybridization of probes to the polytene chromosomes confirmed their expected position on the neo-sex chromosomes of *D. miranda* (their positions on the polytene map¹¹ are: *CycB*, 32A; *robo*, 33C; *eng*, 31D; *eve*, 23A). For *CycB*, *robo* and *eve*, both coding and non-coding sequence data were obtained; for *eng*, only the intron sequence was investigated. The use of reverse-transcriptase-mediated polymerase chain reaction (RT-PCR) with primers spanning an intron confirmed that, for *CycB*, *eve* and *robo*, both alleles are transcribed and spliced in male flies, indicating that both the neo-X and the neo-Y copies are probably under selective constraints. For outgroup comparisons, all genes were isolated by PCR from *D. pseudoobscura*, in which Muller's element C is autosomal (Fig. 1).

We investigated nucleotide variation in 12 wild-derived lines of *D. miranda* (Table 1). A total of 5.2 kilobases per individual were surveyed for the homologous regions on the neo-X and neo-Y chromosomes. Between the neo-X-linked and the neo-Y-linked genes investigated, there are 106 fixed differences and no shared polymorphisms, which is consistent with a lack of recombination between the neo-sex chromosomes. Net silent-site divergence¹² between the neo-X and neo-Y alleles is 2.8%; combining this with previous data¹³, and assuming a silent substitution rate of 1.2×10^{-8} per site per year for *Drosophila*¹⁴, this yields a total divergence time of 2.2 million years, giving an age of about 1.1 million years for the neo-sex chromosome system. On the neo-X chromosome, 61

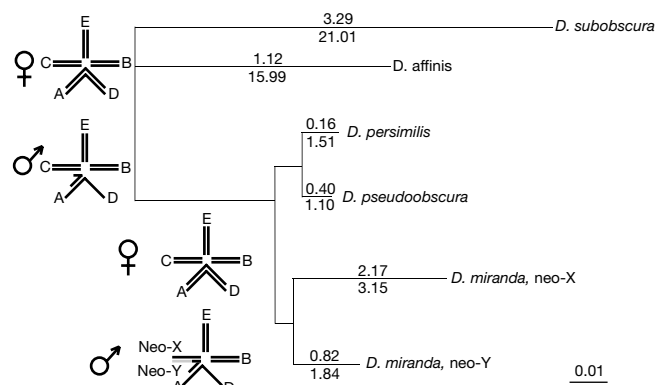


Figure 1 Phylogenetic relationships between the species investigated, constructed from the coding region of *CycB*. The numbers shown above the lines give the estimated K_a values (percentages) for each lineage obtained by maximum likelihood¹⁹; the corresponding K_s values are given below the lines. The karyotypes of the species are also illustrated. The letters A–E indicate the five major chromosomes of the basic *Drosophila* karyotype⁶. In *D. miranda*, element C is fused to the true Y chromosome (neo-Y, shown in grey), and its unfused homologue, the neo-X, segregates with the X chromosome. The relatives of *D. miranda* lack the Y–C fusion. (Note that *D. subobscura* lacks the A–D fusion found in the other species⁶).

Table 1 Patterns of polymorphism and divergence from *D. pseudoobscura* for the neo-sex-linked genes of *D. miranda*

	<i>CycB</i>		<i>robo</i>		<i>eve</i>		<i>eng</i>	
	Neo-X	Neo-Y	Neo-X	Neo-Y	Neo-X	Neo-Y	Neo-X	Neo-Y
Size (base pairs)	1,899	1,827	1,766	1,766	1,228	1,212	387	363
Segregating sites	5	1	20	0	27	1	9	0
θ_W per site (%) ^a	0.212	0.046	0.705	0	1.160	0.045	0.808	0
π per site (%) ^a	0.205	0.023	0.776	0	1.117	0.023	0.747	0
Codons	542	542	1,398	1,398	362	362	0	0
K_S (%)	6.26	5.73	3.16	4.33	4.19	7.05	—	—
K_A (%)	2.51	1.10	0.25	0.93	0.00	0.24	—	—
K_A/K_S	0.40	0.17	0.08	0.21	0.00	0.03	—	—

^a θ_W and π are the two commonly used measures of nucleotide diversity, based on the numbers of segregating sites and mean numbers of pairwise differences between alleles, respectively²⁸; silent-site diversities per site are given.

segregating sites were observed (27 silent single-nucleotide polymorphisms, 4 replacements, 25 non-coding sites and 5 insertion or deletion events). In contrast, the homologous neo-Y-linked regions almost completely lacked variability: only two singleton variants (one synonymous, one non-coding) were detected. Under the standard neutral model, genetic diversity is proportional to the product of the effective population size, N_e , and the mutation rate¹⁵. If no selective forces are operating, N_e for the neo-Y chromosome should be one-third that of the neo-X chromosome, assuming random variation in offspring number for each sex¹⁶. Our data show that the diversity level for the neo-Y chromosome is reduced 30-fold compared with that for the neo-X. This result is consistent with our previous finding of a ~25-fold reduction in microsatellite variability on the neo-Y chromosome of *D. miranda*¹⁷. The overall rate of silent substitution on the neo-Y branch of the tree connecting the neo-X and neo-Y of *D. miranda* and *D. pseudoobscura* (Fig. 1) is similar to that for the neo-X branch, showing that the difference in diversity is not caused by a lower mutation rate of the neo-Y-linked genes (Table 1). A likelihood method was used to quantify the reduction in diversity of the neo-Y chromosome, combining both the microsatellite and the sequence data. Coalescent simulations yield a maximum-likelihood estimate of the ratio of N_e for the neo-Y chromosome to that for the neo-X of 0.04 (Fig. 2). This is consistent with the expectation of a substantial reduction in N_e under various forms of selection acting on a non-recombining genome¹. A similar magnitude of reduction in nucleotide polymorphism has also been observed on an evolving plant Y chromosome¹⁸.

A large reduction in N_e for the neo-Y chromosome greatly reduces the rate of fixation of beneficial mutations on the neo-Y chromosome³, and increases the rate of fixation of deleterious mutations¹. If loci on the neo-X chromosome continue to adapt and their homologues on the neo-Y fail to do so, it is advantageous to upregulate X-linked genes and downregulate their maladapted Y-linked homologues³. Similarly, an accumulation of deleterious alleles on the neo-Y favours increased activity of neo-X genes relative to their neo-Y homologues¹. Both processes can therefore lead to the observed inactivation of neo-Y-linked genes⁹ and to dosage compensation².

Comparisons with the outgroup species reveal that the neo-X-linked copy of *CycB* has a very high rate of protein evolution (Table 1 and Fig. 1). Twenty-three amino-acid replacement substitutions were observed on the branch of the phylogeny leading to the neo-X, but only nine occurred on the neo-Y branch. This difference is not

the result of an increased mutation rate on the neo-X, as the proportion of synonymous substitutions per synonymous site, K_S , is very similar on the neo-X and neo-Y lineages ($K_S \approx 0.06$ for both branches (Table 1)). For phylogenetic analysis of the nucleotide substitutions, *CycB* was also amplified from *D. persimilis*, *D. affinis* and *D. subobscura*, which are all members of the *obscura* subgroup of *Drosophila*, to which *D. miranda* belongs⁶. Likelihood analysis¹⁹ indicates that the ratio of the non-synonymous to synonymous substitution rate (K_A/K_S) differs significantly between branches ($2\Delta L = 31.9$, d.f. = 8, $P < 0.0001$ (Fig. 1)). A model of a single K_A/K_S ratio for all branches was compared with a model of a specific ratio for the neo-X branch plus a second rate common to all other branches (two-rate model). The latter fitted the data significantly better ($2\Delta L = 18.4$, d.f. = 1, $P < 0.00002$), whereas a model with a different K_A/K_S ratio for each branch did not significantly improve the likelihood over the two-rate model ($2\Delta L = 13.4$, d.f. = 7, $P \approx 0.06$). This suggests that most of the heterogeneity in the K_A/K_S ratio between branches of the phylogeny reflects an elevated rate on the branch leading to the neo-X chromosome (Fig. 1).

Heterogeneity in the K_A/K_S ratio between lineages might be caused either by positive selection or by relaxed selective constraints in some lineages¹². If the elevated rate of amino-acid replacements on the neo-X were solely a consequence of reduced functional constraints, the ratio of synonymous to replacement substitutions would be similar for polymorphisms within *D. miranda* and for fixed differences between *D. miranda* and its relatives²⁰. However, a significant excess of the ratio of replacement to silent substitutions relative to polymorphisms was found for comparisons involving the neo-X-linked copy of *CycB* (Table 2), suggesting that directional darwinian selection has driven the rapid evolution of *CycB* on the neo-X chromosome²⁰. In addition, the recent fixation of an advantageous mutation reduces neutral polymorphism around the site that is the target of selection²¹. Consistent with the selection hypothesis is a decrease in variability at *CycB* (Table 1). The level of polymorphism is about 4-fold lower than estimates for the other neo-X-linked genes investigated (see Table 1). By the HKA test²², the level of neo-X silent polymorphism relative to divergence from *D. pseudoobscura* at *CycB* is significantly lower than that for all three other genes (*eve*, *eng* and *robo* ($P < 0.01$)). This result is consistent with the idea that the neo-X chromosome experiences a faster rate of adaptive evolution than the non-recombining neo-Y.

In contrast, *eve* and *robo* both show a higher rate of amino-acid replacements on the neo-Y chromosome than their neo-X homologues (Table 1). Maximum-likelihood analysis of the combined

Table 2 McDonald–Kreitman tests of neutrality at *CycB*

Comparison	Fixed		Polymorphic		P^*
	Synonymous	Replacement	Synonymous	Replacement	
Neo-X/neo-Y	14	30	5	0	<0.01
Neo-X/ <i>D. pseudoobscura</i>	20	27	4	0	0.04
Neo-Y/ <i>D. pseudoobscura</i>	17	12	1	0	0.68

* The P values are from two-tailed Fisher's exact tests, comparing numbers of synonymous versus replacement changes in the fixed and polymorphic categories, respectively.

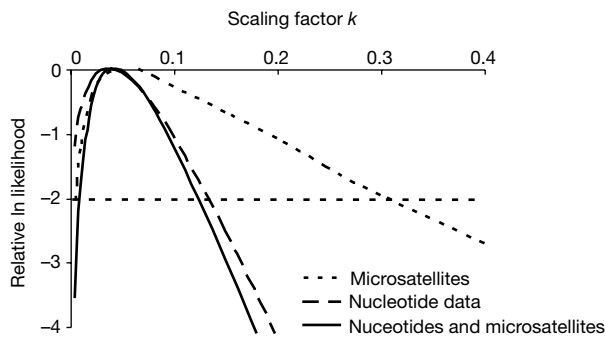


Figure 2 Plots of the relative values of the log-likelihood functions $\ln L(k|\Delta V)$, $\ln L(k|\Delta S)$ and $\ln L(k|\Delta V, \Delta S)$ as a function of k , the reduction in the effective population size of the neo-Y chromosome relative to the neo-X. ΔS and ΔV are the observed differences in numbers of segregating sites and variances of microsatellite repeat numbers between the neo-X and the neo-Y chromosome. All three curves have been normalized so that their maxima are at 0. The horizontal line indicates the two-unit support interval.

sequence data set for *eve* and *robo* was used to compare a model in which K_a/K_s was assumed to be the same on the neo-X and neo-Y branches with a model in which it was allowed to differ; the latter fitted the data significantly better ($2\Delta L = 6.1$, d.f. = 1, $P = 0.01$). In addition, K_a/K_s on the neo-Y branch of *CycB* was higher than on the other branches of the phylogeny (except for the neo-X branch; see Fig. 1). Several lines of evidence suggest that the acceleration of the rate of amino acid substitution on the neo-Y chromosome is not due to a lack of functional constraints on the coding region of *eve* and *robo*. Using RT-PCR, we demonstrated that the neo-Y-linked copies of both genes are transcribed. The entire coding sequences of *eve* and *robo* were analysed, and contained no stop codons or frameshift mutations. A ratio of K_a/K_s of 1 is expected for a gene that is evolving entirely neutrally, as was found for the *Lcp* genes on the neo-Y chromosome of *D. miranda*¹³, whereas a lower ratio is expected for genes subject to functional constraints¹². *Eve* and *robo* display a very low K_a/K_s ratio on the neo-Y chromosome (Table 1), suggesting that both genes are under selective constraints. These results, together with the large reduction in N_e for the neo-Y chromosome, suggest that many amino-acid substitutions in these genes are slightly deleterious and that natural selection has not been able to prevent their accumulation on the neo-Y branch¹. A similar effect has been reported for some other non-recombining genomes^{23,24}.

The adaptive significance of sexual recombination is one of the most puzzling problems in evolutionary biology^{25,26}. Most asexual lineages of eukaryotes seem to become extinct at a higher rate than their sexual relatives^{25,27}. Similarly, the Y chromosome is prone to degeneration once it stops recombining with the X¹. The dismal fates of clonally transmitted species and genomes suggest that genetic recombination is necessary for their persistence over long periods of evolutionary time. Theoretical models predict that sexual populations should be more effective in incorporating new beneficial mutations and preventing the accumulation of deleterious alleles^{25,26}. Our evidence for both faster adaptation on the recombining X chromosome and the accumulation of deleterious mutations on the non-recombining Y chromosome suggests that both processes may be important in conferring an evolutionary advantage to sex and recombination. □

Methods

Isolation of genes, sequencing and RT-PCR

CycB and *robo* were isolated from a genomic library constructed from *D. miranda* males (strain 0101.3 (ref. 13) using the Lambda FixII Library kit (Stratagene). *eve* and *eng* were isolated by using degenerate primers designed from regions conserved between *D. melanogaster* and *D. virilis*. Allele-specific primers were used for PCR amplification of the neo-X-linked and neo-Y-linked copies of *CycB* and *robo* from male genomic DNA, followed by direct sequencing of both strands of the PCR products. The neo-Y-specific

primers only amplify a product in males, while the neo-X-specific primers amplify in both sexes. For *eng* and *eve*, PCR primers amplifying both copies were used, and the amplified product was cloned with the TOPO TA cloning kit (Invitrogen). Clones containing either the neo-X or neo-Y copy were distinguished by a size difference in the cloned PCR products in the case of *eng*, or by a *Hae*II restriction site for *eve*. To exclude PCR errors, at least three independent clones per allele and individual were analysed (the PCR error rate was found to be 0.00092 per base pair). For the population survey, 12 strains of *D. miranda* were sequenced (see ref. 13 for description of the strains), using the ABI Prism BigDye Chemistry (Perkin-Elmer) on an ABI 377 automated sequencer.

Total RNA from male *D. miranda* (strain MSH 38 (ref. 13)) was extracted with the RNA STAT-60 kit, and first-strand cDNA synthesis was performed with random or dT primers and the cDNA Cycle kit (Invitrogen). Primers spanning at least one intron were used to PCR-amplify the neo-X and the neo-Y copies. Amplification products were cloned with the TOPO TA cloning vector and sequenced.

Sequence analysis

Sequences of *D. miranda* and *D. pseudoobscura* were aligned manually. Nucleotide diversity was calculated as described²⁸. To align the coding region of *CycB* between the more diverged members of the *obscura* species group, the translated proteins were aligned by using CLUSTAL X (<http://ftp-igbmc.u-strasbg.fr/pub/ClustalX/>) and the nucleotide alignment was constructed from the protein alignment. The phylogeny based on the coding region of *CycB* was derived by using PUZZLE (<http://ftp.ebi.ac.uk/pub/software/mac/puzzle/>) and is shown in Fig. 1. K_a and K_s were calculated with a maximum-likelihood method, which accounts for unequal transition and transversion rates and unequal base and codon frequencies¹⁹. A likelihood ratio test was used to test different models of evolution by application of the codeml program within the PAML software package¹⁹. A model assuming the same K_a/K_s ratio for all branches was compared with a model assuming a different K_a/K_s ratio for all or some branches of the phylogeny. Twice the difference in log-likelihood between models is assumed to be distributed approximately as χ^2 , with degrees of freedom equal to the difference in the numbers of parameters.

Coalescent process simulations

Coalescent simulations with the standard algorithm of Hudson²⁹ were performed to obtain a maximum-likelihood estimate of the reduction in N_e for the neo-Y chromosome compared with that of the neo-X, combining both the nucleotide polymorphism data set presented here and data on microsatellite variability¹⁷. Whereas all loci on the neo-Y share one genealogy, the loci on the neo-X are probably independent. We therefore generated 11 independent trees for a data set of 12 chromosomes for the neo-X (7 microsatellite loci and 4 genes), whereas a single tree with 11 completely linked loci was computed for the neo-Y. Microsatellite mutations were superimposed on the trees following the single stepwise mutation model described in ref. 17. In short, seven random estimates of $\theta = 4N_e\mu$ (where μ is the mutation rate per microsatellite locus) were drawn from a gamma distribution for each run, and mutations using the resulting θ values were laid down on the trees in accordance with Poisson distributions, using the same values of θ for homologous loci. The mean variance in repeat number per locus, $\bar{V}_i$, across the seven loci was then calculated. After each run, $\Delta V_{sim} = (\bar{V}_X - \bar{V}_Y)/\bar{V}_X$ for the simulated data set was computed. For each gene, the total number of segregating mutations (*CycB*, 6; *robo*, 20; *eve*, 28; *eng*, 9) were assigned to the simulated neo-X and neo-Y trees in proportion to their lengths. After each run, the total numbers of segregating sites on the neo-X (S_X) and on the neo-Y (S_Y) were determined and the quantity $\Delta S_{sim} = (S_X - S_Y)/S_X$ was computed. To estimate the reduction in N_e for the neo-Y loci, the tree length (in units of $2N_e$ generations) was multiplied by a scaling factor k before mutations were laid down on the neo-Y tree. The simulations can be used to determine the likelihood of a pair of $(\Delta V_{sim}, \Delta S_{sim})$ for a given value of k . n replicas are generated for each value of k . The number of runs M_δ where $|\Delta V_{sim} - \Delta V_{obs}| \leq \delta$ and $|\Delta S_{sim} - \Delta S_{obs}| \leq \delta$ was determined, where δ is a preassigned mesh size for the continuous variables ΔV and ΔS . Following ref. 30, the likelihood of the sample is approximated by

$$L(k|\Delta V, \Delta S) = \frac{1}{n} M_\delta(\Delta V, \Delta S)$$

Here, n is 10^5 and δ is 0.02.

Received 21 September; accepted 18 December 2001.

- Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **55**, 1563–1572 (2000).
- Marin, L., Siegal, M. L. & Baker, B. S. The evolution of dosage-compensation mechanisms. *BioEssays* **22**, 1106–1114 (2000).
- Orr, H. A. & Kim, Y. An adaptive hypothesis for the evolution of the Y chromosome. *Genetics* **150**, 1693–1698 (1998).
- Lahn, B. T., Pearson, N. M. & Jégalian, K. The human Y chromosome, in the light of evolution. *Nature Rev. Genet.* **2**, 207–216 (2001).
- Carvalho, A. B., Lazzaro, B. P. & Clark, A. G. Y chromosomal fertility factors *kl-2* and *kl-3* of *Drosophila melanogaster* encode dynein heavy chain polypeptides. *Proc. Natl Acad. Sci. USA* **97**, 13239–13244 (2000).
- Powell, J. R. *Progress and Prospects in Evolutionary Biology: The Drosophila Model* (Oxford Univ. Press, New York, 1997).
- Schaeffer, S. W. & Miller, E. L. Molecular population genetics of an electrophoretically monomorphic protein in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. *Genetics* **132**, 163–178 (1992).
- Gethmann, R. C. Crossing over in males of higher Diptera (Brachycera). *J. Hered.* **79**, 344–350 (1988).
- Steinemann, M. & Steinemann, S. Enigma of Y chromosome degeneration: neo-Y and neo-X

- chromosomes of *Drosophila miranda* a model for sex chromosome evolution. *Genetica* **102**–103, 409–420 (1998).
10. Waters, P. D., Duffy, B., Frost, C. J., Delbridge, M. L. & Graves, J. A. The human Y chromosome derives largely from a single autosomal region added to the sex chromosomes 80–130 million years ago. *Cytogenet. Cell. Genet.* **92**, 74–79 (2001).
 11. Das, M., Mutsuddi, D., Duttagupta, A. K. & Mukherjee, A. S. Segmental heterogeneity in replication and transcription of the X₂ chromosome of *Drosophila miranda* and conservativeness in the evolution of dosage compensation. *Chromosoma* **87**, 373–388 (1982).
 12. Li, W. *Molecular Evolution* (Sinauer Associates, Sunderland, Massachusetts, 1997).
 13. Yi, S. & Charlesworth, B. Contrasting patterns of molecular evolution of the genes on the new and old sex chromosomes of *Drosophila miranda*. *Mol. Biol. Evol.* **17**, 703–717 (2000).
 14. Wang, R. L. & Hey, J. The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the *period* locus. *Genetics* **144**, 1113–1126 (1996).
 15. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
 16. Wright, S. *Evolution and the Genetics of Populations* (Univ. of Chicago Press, Chicago, 1969).
 17. Bachtrog, D. & Charlesworth, B. Reduced levels of microsatellite variability on the neo-Y chromosome of *Drosophila miranda*. *Curr. Biol.* **10**, 1025–1031 (2000).
 18. Filatov, D. A., Moneger, F., Negruțiu, I. & Charlesworth, D. Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404**, 388–390 (2000).
 19. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
 20. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
 21. Barton, N. H. Genetic hitchhiking. *Phil. Trans. R. Soc. Lond. B* **355**, 1553–1562 (2000).
 22. Hudson, R. R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
 23. Lynch, M. & Blanchard, J. L. Deleterious mutation accumulation in organelle genomes. *Genetica* **102**–103, 29–39 (1998).
 24. Fridolfsson, A. K. & Ellegren, H. Molecular evolution of the avian *CHD1* genes on the Z and W sex chromosomes. *Genetics* **155**, 1903–1912 (2000).
 25. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, Cambridge, 1978).
 26. Barton, N. H. & Charlesworth, B. Why sex and recombination? *Science* **281**, 1986–1990 (1998).
 27. Bell, G. *The Masterpiece of Nature* (Univ. of California, Berkeley, 1982).
 28. Tajima, F. Statistical analysis of DNA polymorphism. *Jpn J. Genet.* **68**, 567–595 (1993).
 29. Hudson, R. R. in *Oxford Surveys in Evolutionary Biology* Vol. 7 (eds Futuyma, D. & Antonovics, J.) 1–44 (Oxford Univ. Press, Oxford, 1990).
 30. Weiss, G. & von Haeseler, A. Inference of population history using a likelihood approach. *Genetics* **149**, 1539–1546 (1998).

Acknowledgements

We thank P. Andolfatto, N. Barton, D. Charlesworth, I. Gordo, P. Keightley and S. Wright for helpful comments on the manuscript. D.B. is supported by a Marie Curie fellowship and B.C. by the Royal Society.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.B. (e-mail: doris.bachtrog@ed.ac.uk).

Dissecting the architecture of a quantitative trait locus in yeast

Lars M. Steinmetz^{*†}, Himanshu Sinha[‡], Dan R. Richards^{*}, Jamie I. Spiegelman[†], Peter J. Oefner^{†§}, John H. McCusker[‡] & Ronald W. Davis^{*†§}

^{*}Departments of ^{*}Genetics and [§]Biochemistry, Stanford University School of Medicine, Stanford, California 94305, USA

[†]Stanford Genome Technology Center, 855 California Avenue, Palo Alto, California 94304, USA

[‡]Department of Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA

Most phenotypic diversity in natural populations is characterized by differences in degree rather than in kind. Identification of the actual genes underlying these quantitative traits has proved difficult^{1–5}. As a result, little is known about their genetic architecture. The failures are thought to be due to the different contributions of many underlying genes to the phenotype and

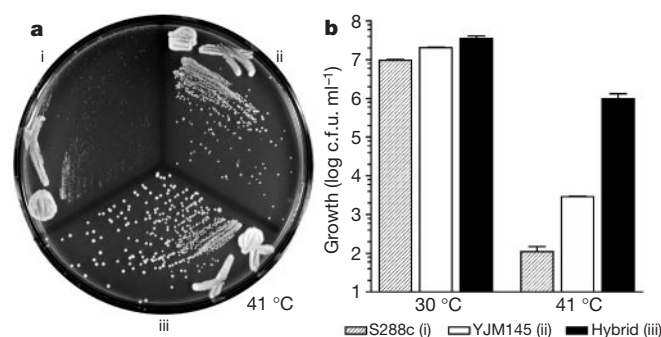


Figure 1 Analysis of the high-temperature-growth phenotype (Htg). **a**, Qualitative differences measured by colony size; **b**, quantitative differences in growth measured by competition assay after 48 h at 30 and 41 °C: (i) S1029, an Htg[−] S288c strain, (ii) YAG040, an Htg⁺ YJM145 strain, and (iii) XHS123, a YJM145/S288c hybrid. Bars indicate s.e.m. ($n = 6$).

the ability of different combinations of genes and environmental factors to produce similar phenotypes^{6,7}. This study combined genome-wide mapping and a new genetic technique named reciprocal-hemizygosity analysis to achieve the complete dissection of a quantitative trait locus (QTL) in *Saccharomyces cerevisiae*. A QTL architecture was uncovered that was more complex than expected. Functional linkages both *in cis* and *in trans* were found between three tightly linked quantitative trait genes that are neither necessary nor sufficient in isolation. This arrangement of alleles explains heterosis (hybrid vigour), the increased fitness of the heterozygote compared with homozygotes. It also demonstrates a deficiency in current approaches to QTL dissection with implications extending to traits in other organisms, including human genetic diseases.

We chose the high-temperature-growth phenotype (Htg), common to clinically derived yeast isolates^{8–10} as a quantitative trait, because this ability is correlated with virulence in studies of infected mice¹¹. We selected two *S. cerevisiae* genetic backgrounds for study. YJM145 is a homozygous diploid strain⁸ that was generated from a heterozygous clinical isolate obtained from the lung of an AIDS patient¹². In contrast, S288c, a commonly used haploid laboratory genetic background for which the whole genome sequence is known, was derived from a strain isolated from a rotten fig¹³.

The respective Htg phenotypes of YJM145, S288c and a diploid hybrid of the two strains (designated YJM145/S288c) were determined by using a colony-size assay (Fig. 1a) and a quantitative competition assay (Fig. 1b). Both assays showed that a diploidized S288c strain grows poorly at high temperature (Htg[−]) relative to YJM145 (Htg⁺). In addition, both assays demonstrated that the hybrid displays heterosis. These data suggested that Htg is co-dominant and that both the clinical and the laboratory genetic backgrounds contain alleles that contribute to the Htg⁺ phenotype of the hybrid.

To analyse the heritability of the Htg phenotype, we sporulated a hybrid that was generated by crossing YJM789, a haploid strain isogenic with YJM145, with S96, a haploid strain isogenic with S288c. Each haploid progeny (segregant) was tested for growth at 41 °C with the use of the colony size assay. The segregants exhibited a wide range of Htg phenotypes, but, interestingly, none of the segregants were as Htg⁺ as the hybrid. Only 104 of 960 segregants (240 tetrads) were at least as Htg⁺ as YJM789 and were selected for further analysis. At this Htg⁺ threshold, defining Htg⁺, the 1:9 inheritance ratio predicted at least 3.2 underlying genetic loci ($1/9 = 1/2^{3.2}$). However, because Htg is quantitative and could be conditioned by combinations of alleles that are each neither

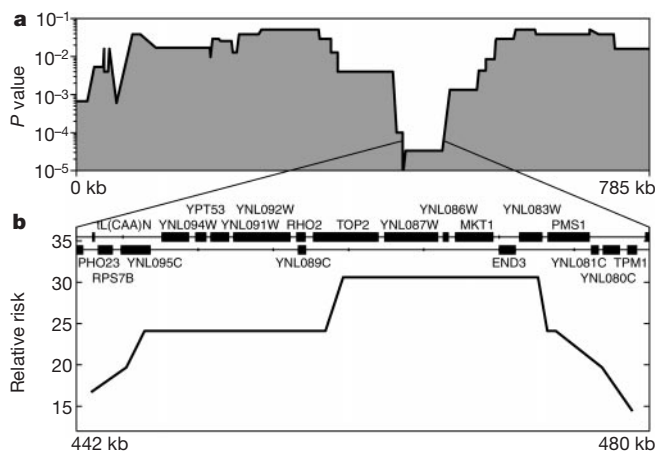


Figure 2 Detection by genome-wide mapping and detailed analysis by fine-structure mapping of the chromosome XIV Htg QTL. **a**, Calculated probability assuming random segregation (y -axis) for 3,444 markers in a whole-genome allelic variation scan of 19 Htg⁺ segregants, shown here for chromosome XIV. A region with low probability was identified between bp 434,194 and 485,856. **b**, Relative-risk plot representing the fine-structure map of the low-probability region. From 104 Htg⁺ segregants, the recombination breakpoints defining an interval with high relative risk were placed to within 32 kb (bp 445,003–477,059 of chromosome XIV).

necessary nor sufficient, the prediction of an exact number of loci is impossible.

A QTL mapping strategy was devised to detect the major Htg⁺ contributors from either genetic background. To identify DNA markers, total genomic DNAs from YJM789 and S96 were hybridized separately to Affymetrix high-density oligonucleotide arrays containing several probes to every open reading frame (ORF) of the yeast genome^{14,15}. A total of 3,444 biallelic markers were identified from probes with decreased signal strength in YJM789 relative to S96–DNA hybridizations, yielding an average marker spacing over the entire genome of 3,504 base pairs (bp), corresponding to an average genetic distance of 1.2 centimorgans (cM).

To maximize the ability to detect QTLs by genotyping the extreme phenotype progeny¹⁶, 19 of the selected 104 Htg⁺ segregants were analysed in genome-wide scans. Genomic DNA from each Htg⁺ segregant was hybridized to a high-density oligonucleotide array. The most probable parental origin for each DNA interval was determined from the hybridization signal. Meiotic recombination breakpoints were identified and compared for all Htg⁺ segregants, to find genomic intervals inherited from the same parent in a significant number of segregants. Two intervals with a low probability

(P) of random segregation were identified: one on chromosome XIV (51.6 kilobases (kb) in size, $P \approx 1/30,000$) and one on chromosome XVI (8.1 kb, $P \approx 1/20,000$).

To confirm the heritability of both intervals and to define precisely the QTL regions, the inheritance of multiple markers in all 104 Htg⁺ segregants was examined. With the use of denaturing high-performance liquid chromatography (DHPLC), 21 markers in the chromosome XVI interval and 28 markers in the chromosome XIV interval were identified and genotyped. The chromosome XVI QTL had a low but significant level of Htg⁺ association: of 104 Htg⁺ segregants, 66.7% inherited YJM145-derived alleles in an interval between 298,521 and 301,584 bp. This percentage translated to a relative risk of 2.1, namely the probability of being Htg⁺ is increased 2.1-fold if a strain carries the YJM145-derived rather than the S288c-derived alleles for this interval. Given its low relative risk, the chromosome XVI interval was not examined further.

A far greater Htg⁺ association was identified for the chromosome XIV QTL: 96.2% of the Htg⁺ segregants inherited the YJM145-derived alleles, yielding a relative risk of 30.6. A peak of 14.9 kb was mapped inside a larger interval of 32 kb (Fig. 2). Interestingly, the interval size was larger than the expected ~ 6 kb or 2 cM, based on the almost 100% Htg⁺ association (200 cM/104 segregants)¹⁷.

Two lines of evidence further strengthened the association of the chromosome XIV interval with Htg⁺. First, random segregants from the same cross displayed random segregation of the QTL¹⁴. Second, DHPLC analysis of 64 Htg⁺ segregants from a similar cross between an Htg⁻ S288c background strain and an unrelated Htg⁺ clinical strain, isogenic with YJM421⁸, showed 87.5% Htg⁺ association (position 468,445) (L.M.S. and J.H.M., unpublished data). The high but incomplete association of the chromosome XIV interval in both crosses confirmed that it is a major-effect QTL, yet neither necessary nor sufficient to condition Htg⁺ in the segregants.

To identify the phenotypically relevant alleles, we performed a detailed sequence analysis of the chromosome XIV interval. In addition to identifying the differences between YJM145 and S288c-derived alleles, the purpose of this comparative sequence analysis was to test the assumption of marker–trait association, namely that phenotypically relevant Htg⁺ sequence variants would be more prevalent in Htg⁺ strains than in Htg⁻ strains. The entire 32 kb interval (445,003–477,059) was therefore sequenced from six Htg⁺ and seven Htg⁻ strains. All Htg⁺ strains are segregants of clinical isolates (Fig. 3, rows a–f), whereas the Htg⁻ strains consist of common laboratory strains (Fig. 3, rows i and k–m), and segregants of wine (row g), grape (row h) and distillery (row j) strains. In total, 374 sequence variants were identified that define a phylogenetic relationship between the strains (see Supplementary Information).

Between YJM789 and S96, 83 sequence variants (1/385 base pairs) were detected, of which 24 were non-synonymous, including three insertion–deletions in YNL089C (457,139), PMS1 (473,368) and

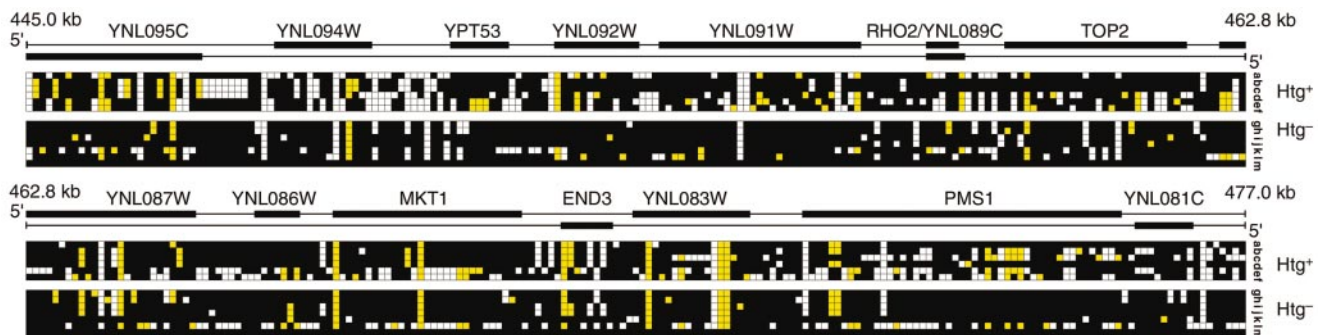


Figure 3 Sequence variations in chromosome XIV QTL between YJM789 (YJM145 genetic background), S96 (S288c background) and five other Htg⁺ and six other Htg⁻ yeast strains. Each column represents a sequence variant relative to S288c; non-

synonymous variants are coloured in yellow. Rows a–f show Htg⁺ strains: YJM789 (a), YJM326 (b), YJM320 (c), YJM280 (d), YJM421 (e), YJM339 (f). Rows g–m show Htg⁻ strains: YJM270 (g), YJM269 (h), YJM627 (i), YJM1129 (j), W303 (k), SK1 (l), S96 (m).

YNL083W (472,573) that encoded two truncated and one extended protein product in YJM789, respectively. Ten non-synonymous variants with identical sequences for all Htg⁺ strains were found in YNL095C, YNL092W, YNL087W, YNL083W and MKT1. However, many of the Htg⁻ strains shared the same sequences, eliminating the significance of the association. Low and incomplete levels of association were also observed for variants in YNL095W, YNL091W and PMS1.

In addition to the limited marker-trait association, none of the genes in the chromosome XIV QTL showed mRNA transcript level differences of threefold or greater in a genome-wide expression analysis of YJM789 and S96 at 30 and 37 °C (see Supplementary Information). Consequently, we developed a new functional assay named reciprocal-hemizygosity analysis to identify the phenotypically relevant allele(s).

Reciprocal-hemizygosity analysis is a tool for analysing each allele

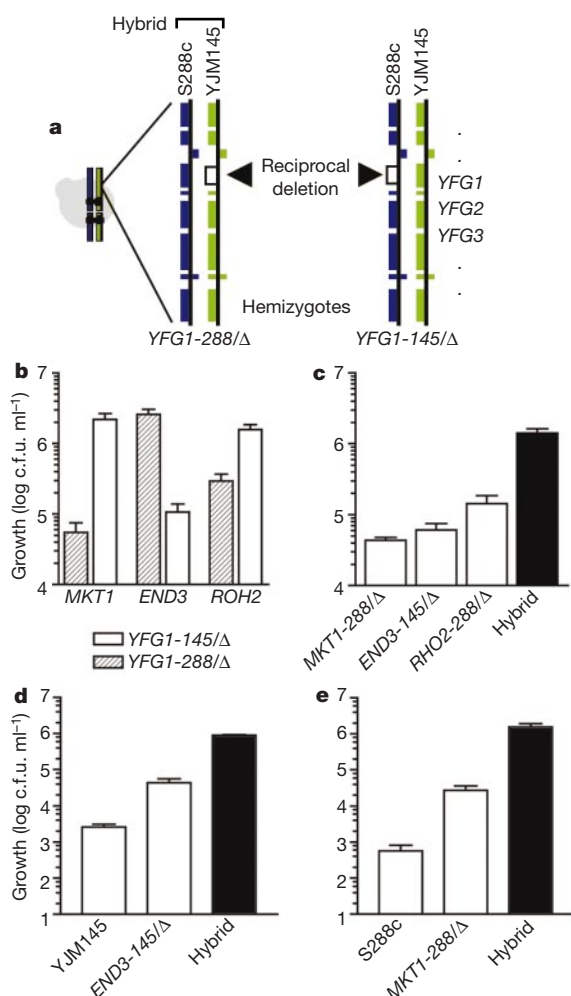


Figure 4 Reciprocal-hemizygosity analysis. **a**, Comparison between reciprocal *YFG1* hemizygotes (a generic gene name). Each block represents a gene. Each hemizygote carries only a single allele of *YFG1* but has an otherwise hybrid diploid genome. Two different dominant drug markers (open boxes) distinguish between competing reciprocal strains. **b**, Reciprocal-hemizygosity analysis of *MKT1*, *END3* and *RHO2* at 41 °C. For each gene pair, 12 measurements ($n = 12$) were taken to calculate the growth mean (c.f.u. ml⁻¹) and s.e.m. **c**, Htg contributions of *MKT1*-145, *END3*-288 and *RHO2*-145 alleles ($n = 12$). **d**, Role of S288c-derived alleles in heterosis. Competitions between YAG040 (Htg⁺ YJM145 background) and the hybrids, XHS122 (*END3*-145/Δ) and XHS123 (YJM145/S288c) ($n = 6$). **e**, Role of YJM145-derived alleles in heterosis. Competitions between S1029 (Htg⁻ S288c background) and the hybrids, XHS119 (*MKT1*-288/Δ) and XHS123 ($n = 6$).

in a hybrid genetic background. We applied reciprocal-hemizygosity analysis to all 15 genes in the chromosome XIV QTL. Isogenic pairs of strains were constructed in the hybrid background (YJM145/S288c) that differed genetically only in the alleles of one gene. In each strain one allele of one gene was deleted, producing a hemizygous diploid carrying either the S288c-derived or the YJM145-derived allele only (Fig. 4a). With reciprocal-hemizygosity analysis, we tested whether an allele from one genetic background was advantageous over that from the other by measuring competitive growth between reciprocal strains.

Because the phenotype is measured in hemizygotes and the identification of QTL alleles relies on a comparison between reciprocal strains, the approach also works for essential genes and is insensitive to potentially confounding gene dosage effects. As measurements are made in the hybrid strain background, segregating alleles from both S288c and YJM145 genetic backgrounds can be detected in one assay. In addition, segregating alleles inactive in each single genetic background but functional in combination with alleles from the other strain background can be discovered. Although our analysis used only single gene deletions, deletion combinations can be used to capture potential synergistic effects contributed by pairs of genes.

All allelic pairs, including the essential gene *TOP2*, were tested at 30, 40 and 41 °C. Strains containing only S288c-derived or YJM145-derived alleles of *YNL095W*, *YNL094W*, *YPT53*, *YNL092W*, *YNL091W*, *TOP2*, *YNL087W*, *YNL086W* and *YNL081C* grew equivalently at all temperatures. Htg differences were also not detected between alleles of *YNL083W* and *PMS1* containing the insertion-deletions. However, the spot dilution experiments showed growth differences for *MKT1*, *END3* and *RHO2*.

RHO2 overlaps with *YNL089C* and these two ORFs cannot be distinguished by deletion analysis. However, it seems that *YNL089C* is not a true ORF: its transcript was not detected in mRNA hybridizations to high-density arrays; three stop codons disrupt the homologous *Candida albicans* sequence; and a recent computational re-annotation of the *S. cerevisiae* genome defined *YNL089C* as a spurious ORF¹⁸. We therefore eliminated *YNL089C* from further consideration.

For *MKT1* and *RHO2*, the YJM145-derived alleles (*MKT1*-145 and *RHO2*-145) were found to confer an Htg⁺ advantage. As determined by the more quantitative plating assay, no significant phenotypic differences were observed between reciprocal hemizygotes after 48 hours at 30 °C. However, at 40 °C the differences were significant. At 41 °C, a 40-fold growth difference was observed between isogenic *MKT1*-145/Δ and *MKT1*-288/Δ hemizygotes and a 5.5-fold difference between the respective *RHO2* hemizygotes ($P < 0.001$) (Fig. 4b).

Surprisingly, for *END3* the S288c-derived allele (*END3*-288) and not the YJM145-derived allele accounted for a 23.6-fold difference

Amino-acid position:	MKT1										END3				RHO2			Htg
	30	106	337	432	453	600	610	657	806		245	258	268		91	150		
YJM789	G	S	F	A	R	F	S	V	A		D	N	N		C	I		+
YJM326	G	S	F	G	R	F	S	V	A		D	N	N		F	I		+
YJM320	G	S	F	G	R	F	S	V	A		D	N	N		F	I		+
YJM280	G	S	F	G	R	F	S	V	A		D	N	N		F	I		+
YJM421	G	S	F	A	R	L	R	I	A		E	S	N		F	I		+
YJM339	G	T	F	A	R	L	R	V	A		D	S	N		F	I		+
YJM270	G	S	F	A	R	F	S	V	A		D	S	N		F	V		-
YJM269	G	S	F	A	R	F	S	V	V		D	S	N		F	V		-
YJM627	G	S	F	A	R	F	S	V	A		D	N	N		F	I		-
YJM1129	G	S	F	A	R	F	S	V	A		D	N	N		F	I		-
W303	G	S	F	A	R	F	S	V	A		D	S	D		C	I		-
SK1	G	S	S	A	R	L	R	V	A		D	S	N		F	I		-
S96	D	S	F	A	K	F	S	V	A		D	S	D		F	I		-

Figure 5 Comparisons of the protein sequence polymorphisms for identified Htg⁺ genes among Htg⁺ and Htg⁻ strains. Amino acid residues are shaded in grey when identical to the Htg⁺ allele sequence. Residues are marked with an asterisk if they are different between the YJM145 and S288c genetic backgrounds.

in growth between isogenic *END3-288/Δ* and *END3-145/Δ* hemizygotes ($P < 0.001$). Although finding an *Htg*⁺ allele from the laboratory-derived strain was not predicted by genetic mapping, the positive contribution of *END3-288* to *Htg* immediately suggested a possible explanation for both the failure to obtain segregants as *Htg*⁺ as the hybrid and for the heterosis seen in crosses between S288c and YJM145 background strains (Fig. 1).

By measuring the relative contribution of each *Htg*⁺ allele in a series of competitions between a hybrid (YJM145/S288c) and, in each case, two hemizygous hybrids that lacked one different *Htg*⁺ allele, the contribution of alleles was found to be greatest for *MKT1-145* and lowest for *RHO2-145* (Fig. 4c). As a control, we confirmed that gene dosage does not affect phenotype: deletion of the *Htg*⁺ alleles (*MKT1-288*, *END3-145* and *RHO2-288*) in hybrid strains and hemizygous deletions of *Htg*⁺ alleles in homozygous YJM145 diploid strains had no detectable phenotypic effect (see Supplementary Information).

The identified *Htg*⁺ alleles are not sufficient to explain the contribution of each genetic background to heterosis. *END3-288* is a significant contributor to the *Htg*⁺ phenotype of the hybrid, as its removal eliminated 48% of the growth advantage of the hybrid over YJM145 (Fig. 4d). However, the remaining *Htg*⁺ phenotype of the *END3-145/Δ* hemizygote suggests the presence of at least one other *Htg*⁺ S288c-derived allele. Removal of *MKT1-145*, the more significant of the two identified YJM145 *Htg*⁺ contributors, eliminated 51% of the growth difference between the hybrid and the laboratory strain (Fig. 4e). Consistent with the already identified *RHO2-145* allele, the remaining *Htg*⁺ phenotype suggests the presence of additional YJM145-derived QTLs.

The fact that neither of the identified alleles is necessary or sufficient is also consistent with a comparison of the protein sequences between *Htg*⁺ and *Htg*[−] strains. There are missense variants in *Htg*⁺ alleles that are common in both populations (Fig. 5) and there are *Htg*[−] strains with *Htg*⁺ protein sequence variants and some with proven *Htg*⁺ functionality (such as *END3-288* in S96) and vice versa. The lack of a single haplotype for all *Htg*⁺ strains indicates that different allele combinations might contribute to *Htg* in different strains. These results are striking in view of the fact that the lack of differences in mRNA levels in these genes, and the lack of gene dosage effects on *Htg*, indicate that the *Htg* phenotype might be caused not by differences in transcription but by changes in protein function. Whereas the function of *Mkt1p* is unknown, apart from its role in the maintenance of the M2 double-stranded RNA virus above 30 °C (ref. 19), *End3p* and *Rho2p* are involved in the cytoskeleton.

Finding three *Htg*⁺ alleles in a single mapped QTL provided immediate explanations for several observations. The two YJM145-derived contributors explain the detection of the QTL, their linkage *in cis* explains the larger than expected map interval, and the identification of the additional S288c-derived contributor *in trans* explains heterosis. The fact that neither single allele is necessary or sufficient for *Htg*⁺ explains the lack of marker–trait association.

These findings indicate that existing approaches to quantitative traits demand re-evaluation. Combinations of both common and rare variants are likely to underlie quantitative traits and the number of genes could be far greater than expected. If closely linked loci are common, current single-gene-per-locus approaches might have intrinsic deficiencies. Although narrowing an interval in the hope of achieving a map interval that approaches a single point might serve to locate the major contributor, the effects of neighbouring genetic factors could be missed. In addition, many QTLs located throughout the genome might not be detected because of their small effect or their location *in trans* with another QTL allele. A more comprehensive approach to QTL dissection therefore needs to include one in which multiple genes can be tested individually and in combination. The reciprocal-hemizygosity analysis has the

potential of being applied genome-wide in model organisms. By using drug resistance markers in combination with molecular barcode tags²⁰, it might be possible to expand our analysis to measuring the phenotypic contribution of all allelic differences between two yeast genomes in a single step. □

Methods

Yeast growth media and strain construction

YPD medium, containing G418 (200 μg ml^{−1}), hygromycin B (300 μg ml^{−1}) or nourseothricin (100 μg ml^{−1}), was used to select strains with *kanMX4*, *hphMX4* and *natMX4* dominant drug-resistance markers^{21,22}. Haploid S288c background strains were diploidized²³. All deletion mutations were constructed in diploids by replacing specific ORFs with dominant drug-resistance cassettes^{21,22}; homologous integration was verified by colony polymerase chain reaction (PCR)²⁴ at both ends of integrated cassettes. Reciprocal hemizygous diploid strains were made with different, phenotypically neutral²¹ dominant drug markers to distinguish between competing strains. As a test of reproducibility, ORFs were replaced with both *hphMX4* and *natMX4* cassettes.

To construct isogenic hemizygous diploids containing only S288c-derived or only YJM145-derived alleles of specific genes, diploid S288c and YJM145 background strains carrying *lys2* or *lys5* mutations, with and without a hemizygous deletion of a gene (see Supplementary Information), were sporulated²⁵. To generate reciprocal pairs, spore suspensions from homozygous S288c and hemizygous YJM145 background diploids were mixed and vice versa. Mixtures were incubated overnight at 30 °C on YPD medium to allow mating between haploid spores, and then replica-plated onto SD medium to select for *Lys*⁺ diploids. *Lys*⁺ diploids were streaked onto the appropriate drug plate to isolate YJM145/S288c diploids hemizygous at a specific gene. See Supplementary Information for a table of strains.

Phenotypic determination of high-temperature growth

Colony size⁸ measured in duplicate was primarily used to determine the *Htg* phenotype of segregants, because *Htg* was more reproducible on plates than in liquid culture. After growth overnight at 30 °C, strains were streaked onto fresh YPD plates to obtain single colonies. After incubation for 48 h at 41 °C, colony size was compared with that of a YJM789 control; measurements above this threshold were defined as *Htg*⁺.

In reciprocal-hemizygosity analysis, one hemizygote carrying only an S288c-derived allele and another carrying only a YJM145-derived allele of a gene were competed, each marked with a different dominant drug marker. Each hemizygote in a pair was grown separately overnight at 30 °C in liquid YPD; cells from both cultures were mixed and diluted to ~10³ cells ml^{−1}. A 1 ml sample of this cell mixture (plus 4 ml sterile water to improve cell dispersal) was placed onto nitrocellulose filter papers 90 mm in diameter and with a pore size of 0.22 μm; the vacuum-dried filters were put on YPD plates and incubated at 30, 40 and 41 °C for 48 h. After incubation, the cells were recovered in 5 ml sterile water and tenfold serial dilutions were spotted onto drug plates. The dilution factor at which no growth was observed was compared for each hemizygote in a pair. For all genes in the *Htg* region, this serial-dilution-spotting assay was performed twice with each of two pairs of hemizygous diploids (dominant drug markers swapped).

In addition to the parental strains in Fig. 1, the *RHO2/Δ*, *MKT1/Δ* and *END3/Δ* hemizygotes, which showed allele-specific *Htg* differences in the dilution-spotting assay, were compared by spreading the dilutions of the suspensions to single colonies onto different drug plates that were incubated at 30 °C for 48 h. After incubation, the numbers of colonies growing on each drug plate were counted to calculate colony-forming units (c.f.u.) ml^{−1}. These competitions were performed twice with each of four pairs of hemizygous diploids to calculate growth means and their standard errors.

Genetic typing and expression analysis

For genome scans, total genomic DNA was fragmented, labelled and hybridized to Affymetrix Ye6100 yeast arrays. Biallelic markers were determined by the decreased hybridization efficiency of YJM789 relative to S96 (ref. 14). Additional markers for fine-structure mapping were identified and genotyped by DHPLC²⁶. Dideoxy sequencing was performed on ~600-bp PCR products tiled over the 32-kb interval with ~50 bp overlaps with the use of big dye terminators. Gene expression analysis was performed on Affymetrix Ye6100 yeast arrays from cultures grown in liquid YPD medium to 2 × 10⁷ cells ml^{−1}; mRNA was processed as described²⁷.

Statistical analysis

The *Htg*⁺ segregant set included pairs from the same tetrad and segregants from unrelated meioses. To increase mapping power, segregant pairs were favoured for the genome-wide scan. The probability of observing a given marker segregation by chance was calculated as a product of multinomial and binomial probabilities that took into account the presence of *Htg*⁺ segregant pairs:

$$p = \left(\frac{n!}{a!b!c!} \right) \left(\frac{1}{6} \right)^a \left(\frac{2}{3} \right)^b \left(\frac{1}{6} \right)^c C(m, k) 2^{-m}$$

given the number of segregant pairs n , of which b have different genotype, a the same YJM789 genotype, and c the same S96 genotype, and given the number of unrelated segregants m , of which k have the YJM789 genotype. Relative risk r was estimated, assuming that the risk for *Htg*⁺ is independent for segregants of a pair, by maximizing the

likelihood function:

$$\left(\frac{r^2}{r^2 + 4r + 1}\right)^a \left(\frac{4r}{r^2 + 4r + 1}\right)^b \left(\frac{1}{r^2 + 4r + 1}\right)^c \left(\frac{r}{r+1}\right)^k \left(\frac{1}{r+1}\right)^{m-k}$$

where r represents the ratio of Htg⁺ probability given YJM145 genotype over that given S288c; other variables as above. For Htg competitions, t -tests were performed to determine the significance of differences in means. Percentage growth deficiency was calculated as $100(\log C - \log B)/(\log C - \log A)$, where A corresponds to c.f.u. ml⁻¹ of the control, B to that of the hemizygote, and C to that of the hybrid.

Received 9 August 2001; accepted 25 January 2002.

- Lander, E. S. & Schork, N. J. Genetic dissection of complex traits. *Science* **265**, 2037–2048 (1994).
- Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–856 (2000).
- Darvasi, A. Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genet.* **18**, 19–24 (1998).
- Mackay, T. F. Quantitative trait loci in *Drosophila*. *Nature Rev. Genet.* **2**, 11–20 (2001).
- Mauricio, R. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature Rev. Genet.* **2**, 370–381 (2001).
- Flint, J. & Mott, R. Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Rev. Genet.* **2**, 437–445 (2001).
- Schafer, A. J. & Hawkins, J. R. DNA variation and the future of human genetics. *Nature Biotechnol.* **16**, 33–39 (1998).
- McCusker, J. H., Clemons, K. V., Stevens, D. A. & Davis, R. W. Genetic characterization of pathogenic *Saccharomyces cerevisiae* isolates. *Genetics* **136**, 1261–1269 (1994).
- Clemons, K. V., McCusker, J. H., Davis, R. W. & Stevens, D. A. Comparative pathogenesis of clinical and nonclinical isolates of *Saccharomyces cerevisiae*. *J. Infect. Dis.* **169**, 859–867 (1994).
- Murphy, A. & Kavanagh, K. Emergence of *Saccharomyces cerevisiae* as a human pathogen: Implications for biotechnology. *Enzyme Microb. Technol.* **25**, 551–557 (1999).
- McCusker, J. H., Clemons, K. V., Stevens, D. A. & Davis, R. W. *Saccharomyces cerevisiae* virulence phenotype as determined with CD-1 mice is associated with the ability to grow at 42°C and form pseudohyphae. *Infect. Immun.* **62**, 5447–5455 (1994).
- Tawfik, O. W., Papasian, C. J., Dixon, A. Y. & Potter, L. M. *Saccharomyces cerevisiae* pneumonia in a patient with acquired immune deficiency syndrome. *J. Clin. Microbiol.* **27**, 1689–1691 (1989).
- Mortimer, R. K. & Johnston, J. R. Genealogy of principal strains of the yeast genetic stock center. *Genetics* **113**, 35–43 (1986).
- Winzler, E. A. *et al.* Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197 (1998).
- Steinmetz, L. M. & Davis, R. W. High-density arrays and insights into genome function. *Biotechnol. Eng. Rev.* **17**, 109–146 (2000).
- Lander, E. S. & Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199 (1989).
- Boehnke, M. Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes. *Am. J. Hum. Genet.* **55**, 379–390 (1994).
- Wood, V., Rutherford, K. M., Ivins, A., Rajandream, M. A. & Barrell, B. A re-annotation of the *Saccharomyces cerevisiae* genome. *Comp. Funct. Genom.* **2**, 143–154 (2001).
- Wickner, R. B. MKT1, a nonessential *Saccharomyces cerevisiae* gene with a temperature-dependent effect on replication of M2 double-stranded RNA. *J. Bacteriol.* **169**, 4941–4945 (1987).
- Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* **14**, 450–456 (1996).
- Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* **15**, 1541–1553 (1999).
- Wach, A., Brachat, A., Pohlmann, R. & Philippsen, P. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**, 1793–1808 (1994).
- Herskowitz, I. & Jensen, R. E. in *Methods in Enzymology* (eds Guthrie, C. & Fink, G. R.) 132–246 (Academic, San Diego, 1991).
- Niedenthal, R. K., Riles, L., Johnston, M. & Hegemann, J. H. Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast. *Yeast* **12**, 773–786 (1996).
- Codon, A. C., Gasent-Ramirez, J. M. & Benitez, T. Factors which affect the frequency of sporulation and tetrad formation in *Saccharomyces cerevisiae* baker's yeasts. *Appl. Environ. Microbiol.* **61**, 630–638 (1995).
- Spiegelman, J. I. *et al.* Cloning of the *Arabidopsis* RSF1 gene by using a mapping strategy based on high-density DNA arrays and denaturing high-performance liquid chromatography. *Plant Cell* **12**, 2485–2498 (2000).
- Primig, M. *et al.* The core meiotic transcriptome in budding yeasts. *Nature Genet.* **26**, 415–423 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank N. Risch, E. Mignot, K. White, R. Hyman, J. Haber, C. Scharfe, T. Jones and M. Mindrinos for helpful discussion, and M. Trebo for help in preparing the website. This work was supported by the NIH (P.J.O., J.H.M. and R.W.D.) and by a Howard Hughes Medical Institute predoctoral fellowship awarded to L.M.S. Strain phenotyping and reciprocal-hemizygosity analysis were performed at Duke University; mapping, expression and sequence analysis were done at Stanford University.

Correspondence and requests for materials should be addressed to J.H.M. (e-mail: mccus001@mc.duke.edu) or L.M.S. (e-mail: larsms@stanford.edu). Sequences have been deposited in GenBank under the following accession numbers: S96 (AF458969), YJM280 (AF458970), YJM320 (AF458971), YJM326 (AF458972), YJM339 (AF458973), YJM421 (AF458974), YJM789 (AF458975), SK1 (AF458976), W303 (AF458977), YJM1129 (AF458978), YJM269 (AF458979), YJM270 (AF458980), YJM627 (AF458981).

Inhibition of climbing fibres is a signal for the extinction of conditioned eyelid responses

Javier F. Medina, William L. Nores & Michael D. Mauk

W. M. Keck Center for the Neurobiology of Learning and Memory, and Department of Neurobiology and Anatomy, University of Texas Medical School, Houston, Texas 77030, USA

A fundamental tenet of cerebellar learning theories asserts that climbing fibre afferents from the inferior olive provide a teaching signal that promotes the gradual adaptation of movements^{1–3}. Data from several forms of motor learning provide support for this tenet^{4–8}. In pavlovian eyelid conditioning, for example, where a tone is repeatedly paired with a reinforcing unconditioned stimulus like periorbital stimulation, the unconditioned stimulus promotes acquisition of conditioned eyelid responses by activating climbing fibres^{9–12}. Climbing fibre activity elicited by an unconditioned stimulus is inhibited during the expression of conditioned responses^{9–11}—consistent with the inhibitory projection from the cerebellum to inferior olive^{6,13}. Here, we show that inhibition of climbing fibres serves as a teaching signal for extinction, where learning not to respond is signalled by presenting a tone without the unconditioned stimulus. We used reversible infusion of synaptic receptor antagonists to show that blocking inhibitory input to the climbing fibres prevents extinction of the conditioned response, whereas blocking excitatory input induces extinction. These results, combined with analysis of climbing fibre activity in a computer simulation of the cerebellar-olivary system^{14–16}, suggest that transient inhibition of climbing fibres below their background level is the signal that drives extinction.

To examine how inhibitory inputs to the climbing fibres contribute to extinction of the conditioned response, we infused the GABA (γ-aminobutyric acid) antagonist picrotoxin into the contralateral inferior olive of well trained rabbits undergoing extinction. Twenty rabbits were initially included in the study, of which four were found through histological methods to have cannula placements in the correct region of the inferior olive (two of these four placements are shown in the left column of Fig. 1d). Thus all of our figures and findings report data from these four animals. Rabbits were trained initially for five daily sessions (108 tone plus unconditioned stimulus trials per session) until they had acquired robust eyelid responses. Subsequently, each rabbit received a total of three daily extinction sessions (108 tone alone trials per session) with a different treatment each day (no infusion, continuous artificial cerebrospinal fluid (ACSF) infusion, or continuous infusion with picrotoxin (150 μM, 0.1 μl min⁻¹). To maintain high response levels before infusion, each extinction session was preceded by two daily training sessions of the tone plus unconditioned stimulus. With no infusion or infusion with ACSF, the four rabbits showed normal extinction of responses within the 108-trial session (Fig. 1a). In contrast, during intra-olivary infusion of picrotoxin, the same four rabbits maintained robust response levels throughout the entire extinction session (Fig. 1b). A within subjects analysis of variance and subsequent *F*-test for simple effects revealed that the rates of extinction for no infusion and ACSF infusions were indistinguishable ($F_{11,99} = 0.97$, not significant (NS)), and that both were significantly different from the response rate during picrotoxin infusion ($F_{11,99} = 17.22$ (no infusion) and 16.90 (ACSF), $P < 0.001$) (Fig. 1c).

Unlike experimental manipulations that abolish responses, block learning, or cause extinction, this impairment in extinction of the

conditioned response observed with reversible infusion of picrotoxin provides many internal controls, because response is maintained throughout the infusion. For example, the effects cannot be due to blockade of tone or unconditioned stimulus pathways by picrotoxin. The former would have abolished responses, and the latter is irrelevant as the unconditioned stimulus was omitted during the extinction session. Furthermore, infusion of picrotoxin did not block pathways necessary for response expression or cause a nonspecific performance deficit, as both of these would have abolished responses. Although picrotoxin might produce cerebellar malfunction as a consequence of a tonic increase in climbing fibre activity, infusion of picrotoxin into the inferior olive at the same concentration used here has been shown to increase climbing fibre activity only slightly, to approximately 2 Hz (ref. 17), which is less than the 5–10 Hz stimulation required to turn off Purkinje cells¹⁸. Finally, the possibility that the responses seen after infusion of picrotoxin are not conditioned responses can be excluded, because they retained their adaptive (that is, learned) timing throughout the extinction session. These controls, together with the observation that effective cannula placements were restricted to the rostromedial portions of the dorsal accessory olive (the region that is activated by the unconditioned stimulus during eyelid conditioning^{10,12}; Fig. 1d, two out of the four effective placements are shown in the left column), suggest that picrotoxin prevented the extinction of conditioned responses by blocking inhibitory transmission in the inferior olive.

Next, we examined how excitatory input to the climbing fibres contributes to the maintenance and extinction of conditioned eyelid responses. The same four rabbits with cannula placements in the dorsal accessory olive were re-trained after the final extinction

session, and then were given intra-olivary infusions of either ACSF or the α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)-receptor antagonist NBQX during two separate tone plus unconditioned stimulus test sessions. ACSF infusion did not affect the high level of response during the first tone plus unconditioned stimulus test session (Fig. 2a). In contrast, continuous infusion of NBQX (150 μ M, 0.1 μ l min⁻¹) during a subsequent tone plus unconditioned stimulus session resulted in the gradual disappearance of conditioned eyelid responses, until only the reflex response to the reinforcing unconditioned stimulus remained (Fig. 2b). The decline in conditioned response during infusion of NBQX was statistically indistinguishable from that observed during normal tone-alone extinction training in the same four animals ($F_{11,99} = 0.50$, NS) (Fig. 2c, black line).

These data alone, however, cannot distinguish between an extinction-like decline in response that depends on the presentation of tone-alone trials, and a delayed abolition of conditioned responses due to slow diffusion of NBQX to a site necessary for response expression. To distinguish between these possibilities, the same four rabbits were re-trained and re-tested with an additional tone plus unconditional stimulus session that started 40 min after the continuous infusion of NBQX had begun. During the first test session (in which infusion of NBQX was not delayed), eyelid responses had disappeared completely 40 min after the beginning of the infusion (approximately 60 trials, Fig. 2b). Thus, if NBQX were causing a time-dependent effect, no response would be expected at the time when the second NBQX session began. However, all four rabbits

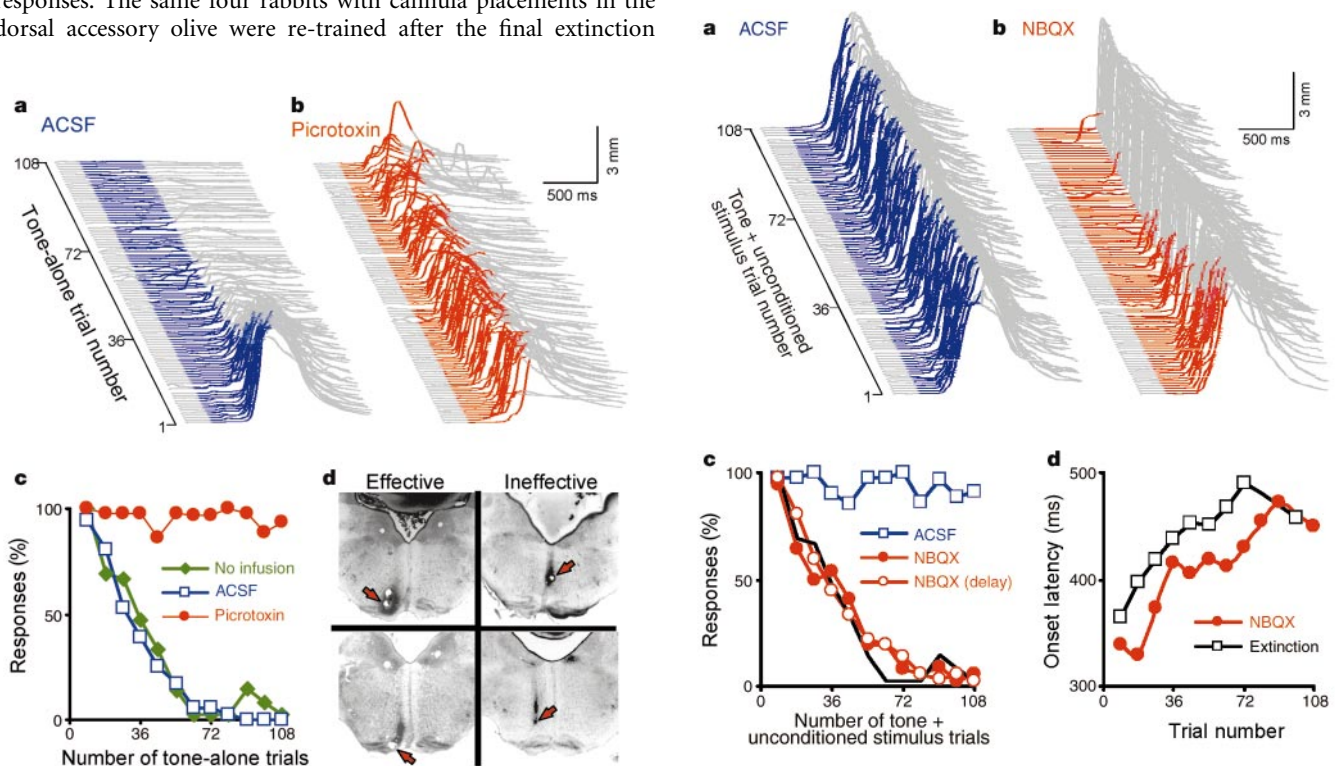


Figure 1 Infusion of picrotoxin into the inferior olive prevented extinction of conditioned responses. **a**, During tone-alone training with infusion of ACSF, conditioned responses gradually extinguished. Each trace represents the average response of the four rabbits for that particular trial. The blue section indicates tone presentation. **b**, With picrotoxin infusion the same four rabbits maintained high response levels throughout the session. The red section indicates tone presentation. **c**, Summary of response percentage for ACSF, picrotoxin, and no infusion sessions for the four rabbits with effective cannula placements. **d**, Four representative cannula placements (arrows) for two of the four animals whose data are shown in **a–c** (left column) and for two of the 16 animals with ineffective cannula placements (right column).

Figure 2 Infusion of NBQX into the inferior olive caused extinction of conditioned responses during tone plus unconditioned stimulus trials. **a**, Normal levels of conditioned response during tone plus unconditioned stimulus sessions were not affected by continuous ACSF infusion into the inferior olive. As in Fig. 1, traces are the average of the four rabbits with successful cannula placement. The blue section indicates tone presentation. **b**, With NBQX infusion, conditioned responses gradually decline until only reflex responses remain. The red section indicates tone presentation. **c**, The rate of response decrement was similar during normal extinction (black line) and during tone plus unconditioned stimulus trials given immediately after or 40 min after infusion of NBQX (see Methods). **d**, Concomitant with the decline in response, NBQX infusion produced a gradual increase in response latency similar to normal extinction.

responded during the initial trials of the second NBQX session, and the subsequent decline of eyelid responses was indistinguishable from the first NBQX test session ($F_{11,99} = 0.34$, NS) (Fig. 2c). NBQX infusion also produced changes in response timing that were indistinguishable from those that occur in extinction. During normal extinction, the latency to onset of the conditioned responses increases progressively as responses disappear gradually (Fig. 2d, black squares). We observed similar changes in onset latency during infusion of NBQX (Fig. 2d, red circles). Together, these findings suggest that intra-olivary infusion of NBQX produced trial-dependent extinction of conditioned responses and not time-dependent abolition.

A previously described computer simulation of the cerebellar-olivary system provides a working hypothesis to interpret the present infusion data^{14–16}. Figure 3a summarizes the connectivity of the simulation that is relevant. Empirical evidence indicates that inhibitory and excitatory inputs to the inferior olive (Fig. 3a) combine with intrinsic cellular properties to regulate climbing fibre activity to approximately 1 Hz (refs 6, 19). We have shown previously that in the simulation, background climbing fibre activity is naturally self-regulated to approximately 1.2 Hz (Fig. 3b, red line), an equilibrium level at which no learning occurs^{14–16}. As such, both acquisition and extinction of the conditioned response require that climbing fibre activity deviate from the equilibrium level during the simulated tone plus unconditioned stimulus or tone-alone trials. Whereas the increases in climbing fibre activity that promote acquisition of the conditioned stimulus

follow fairly intuitively from the activation of excitatory inputs by the unconditioned stimulus (Fig. 3a, unconditioned stimulus input), the simulation suggests factors that may decrease climbing fibre activity below equilibrium to promote extinction of the conditioned response.

The changes in climbing fibre activity over simulated acquisition and extinction training demonstrate an interpretation of the infusion results (red line in Fig. 3c, d). We began by training the simulation with 500 tone plus unconditioned stimulus trials (Fig. 3c). Initially, the unconditioned stimulus engaged the excitatory input to the inferior olive and was able to activate climbing fibres reliably (Fig. 3c, red histogram for trials 0–50) because cerebellar output was low (Fig. 3c, green histogram for trials 0–50) and thus the inhibitory input to the climbing fibres was weak. This reliable activation of climbing fibres by the unconditioned stimulus promoted the acquisition of conditioned responses (indicated by a gradual increase in cerebellar output; Fig. 3c, green line). However, the ability of the unconditioned stimulus to activate climbing fibres decreased as conditioned responses were gradually acquired (Fig. 3c, red line; compare red histograms in the left and right panels). The reason for this decline is that climbing fibres are increasingly inhibited by the increased cerebellar output that generates the conditioned response. This yielded climbing fibre activity near the equilibrium level in well trained simulations, despite the excitatory drive from the unconditioned stimulus (Fig. 3c, red histogram for trials 450–500), consistent with the behaviour of climbing fibres seen in several recording studies^{9–11}.

As recording data have shown that the ability of the unconditioned stimulus to activate climbing fibres is inhibited during a conditioned response^{9–11}, the simulations suggest that a simple extension of this hypothesis may explain the present infusion results: that is, strong inhibition during a conditioned response at a time when the unconditioned stimulus is omitted (as with extinction training) drives climbing fibre activity below its equilibrium level, thereby signalling the induction of plasticity responsible for extinction. Omitting the unconditioned stimulus after acquisition, or blocking the excitatory input to the climbing fibre as with NBQX infusion, tipped the balance in favour of inhibition, and thus promoted extinction by causing climbing fibre activity during the trial to fall well below equilibrium (Fig. 3c, red star labelled NBQX for NBQX infusion; Fig. 3d, red histogram for trials 0–50 for normal extinction). Conversely, although omitting the unconditioned stimulus during extinction training removed the strong excitatory input from the unconditioned stimulus, simulated infusion of picrotoxin also blocked the strong inhibitory input associated with making a conditioned response. Thus, climbing fibre activity did not decrease below the spontaneous level during the trial (Fig. 3d, red star labelled PTX), and extinction was prevented. This interpretation of the effects of picrotoxin may also explain why manipulations that block cerebellar nucleus output directly also prevent extinction²⁰. Thus, our infusion data support the hypothesis that the induction of extinction is signalled by a decrease in climbing fibre activity during the tone, and that this decrease is produced by strong response-associated inhibition occurring in the absence of the excitatory drive from the omitted unconditioned stimulus.

The simulation incorporated a site of plasticity in the cerebellar cortex on the basis of observations that co-activation of granule cells and climbing fibres induces long-term depression (LTD) of granule to Purkinje synapses^{21,22}, whereas granule activity alone induces long-term potentiation (LTP)²². Thus, the bi-directional modulation of climbing fibre activity shown in Fig. 3 was responsible for acquisition and extinction of eyelid responses through the induction of LTD and LTP, respectively. This hypothesis seemingly contradicts evidence supporting the widely held view that extinction does not simply reverse changes that mediate acquisition. The evidence most frequently cited against 'unlearning' mechanisms of extinction is the phenomenon of savings, where relearning after

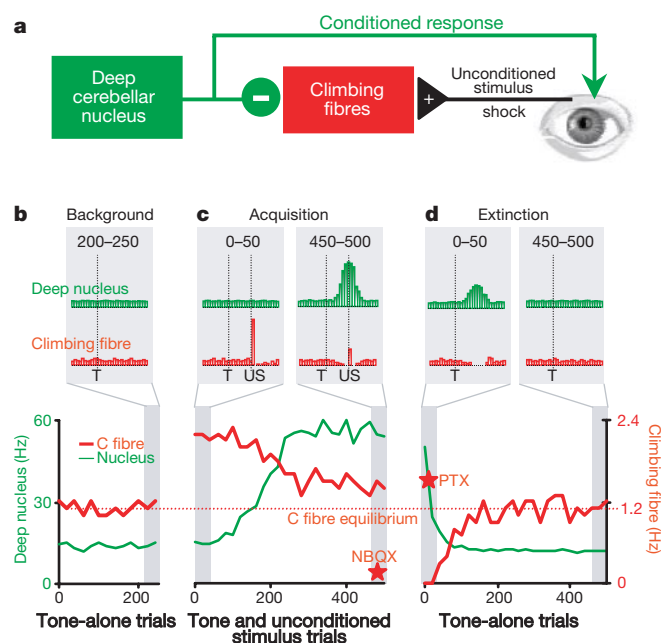


Figure 3 Activity of simulated climbing fibres and cerebellar nucleus cells during acquisition and extinction training. **a**, Schematic representation of climbing fibre afferents. **b–d**, Climbing fibre (red) and nucleus cell (green) activity (charts along the bottom), and activity from the training trials (histograms along the top). T, presentation of the tone; US, presentation of the unconditioned stimulus. **b**, Before training, climbing fibre and nucleus activity are relatively stable. **c**, During acquisition training, climbing fibre activity during the tone plus unconditioned stimulus trials is initially higher owing to the unconditioned stimulus. As the simulation learns, increased nucleus cell activity during the tone inhibits climbing fibre activity to near normal levels. **d**, At the outset of extinction training, climbing fibre activity during presentation of the tone is below normal owing to strong inhibition from the nucleus cells in the absence of excitation from the unconditioned stimulus. NBQX causes extinction by bringing climbing fibre activity below the normal level (star labelled NBQX), whereas picrotoxin blocks extinction by preventing climbing fibre activity from decreasing (star labelled PTX).

extinction is faster than original learning. We have shown recently, however, that a site of plasticity outside of the cerebellar cortex (possibly the cerebellar nuclei) is resistant to extinction and contributes to savings¹⁶. At least for cerebellar learning, therefore, phenomena such as savings are not inconsistent with the idea that acquisition and extinction can involve LTP and LTD at the same set of synapses given that there are at least two sites of plasticity.

The behaviour of climbing fibres during simulated acquisition and extinction is similar to the bi-directional modulation of activity observed in the dopamine neurons of the basal ganglia in response to unexpected reward or to the omission of expected reward^{23,24}. In addition, the predicted contribution of response-associated inhibition in signalling extinction shares important features with the proposed role of feedback inhibition in the regulation of fear conditioning²⁵. Thus, spontaneous activity tightly regulated by feedback inhibition, and the integration of excitatory inputs with feedback inhibition to modulate this activity in both directions, may represent general mechanisms for bi-directional learning. □

Methods

Subjects

We obtained data from 4 male New Zealand albino rabbits (*Oryctolagus cuniculus*), each weighing 2.5–3.0 kg (in addition, 16 rabbits were not included in the study because histological examination revealed that the cannulas were misplaced). The animals were individually housed and given food and water *ad libitum*.

Surgery

All animals were prepared with a cannula implanted in the inferior olive and with a head bolt cemented to the skull. Animals were pre-anaesthetized with 5 mg kg⁻¹ acepromazine, and their skulls were immobilized in a stereotaxic restrainer. Anaesthesia was maintained with isoflurane (2–3% mixed in oxygen), and sterile procedures were used during the placement of the cannulas. The head was positioned with lambda 1.5 mm ventral to bregma. A cannula (Plastics One) consisting of a 26-gauge stainless steel guide sheath was placed at stereotaxic coordinates corresponding to the dorsal accessory olive (from the lambda landmark on the skull: 1 mm anterior, 0.7 mm lateral, 21.9–23.2 mm ventral). After placement, the cannula and head bolt were secured to the skull with dental acrylic. In addition, two stainless steel stimulating electrodes were chronically implanted in the periorbital muscles rostral and caudal to the eye. Treatment of the animals and surgical procedures were in accordance with the National Institutes of Health Guidelines, and were approved by the University of Texas, Houston, animal welfare committee.

Conditioning

Each daily training session consisted of 12 blocks of 9 trials. For paired tone plus unconditioned stimulus sessions, each block consisted of eight paired presentations of the tone and unconditioned stimulus and one presentation of the tone by itself. The tone (1 kHz, 85 dB) was presented for 550 ms during tone-alone trials, and co-terminated with a 50 ms train of constant pulses (200 Hz, 1 ms pulse width, 2–3 mA) delivered to the periorbital electrodes during paired trials. Trials were separated by a random inter-trial interval in the range of 25–35 s. We recorded movement of the unrestrained eyelid by measuring the reflectance of an infrared light-emitting diode aimed at the eyelid.

Data analysis

Digitized sweeps of eyelid movement corresponding to the 200 ms before and 2,300 ms after the tone onset were analysed using custom software. To be counted as a conditioned response, the movement amplitude had to reach 0.3 mm within 500 ms after tone onset. Onset latency was determined by calculating the point at which the response reached the 0.3 mm criterion. Trials in which there was greater than 0.3 mm of movement during the 200 ms baseline period collected before tone onset were excluded from additional analysis.

Infusions

Infusions of ACSF, picrotoxin (150 µM), or NBQX (150 µM) were accomplished with a 33-gauge internal cannula that extended 1.2 mm beyond the tip of the guide cannula. Each infusion started 10 min before the beginning of the conditioning session at a rate of 0.1 µl min⁻¹. Then conditioning began and infusion continued throughout the session at 0.1 µl min⁻¹. In one set of experiments (NBQX delay), the conditioning session started 40 min after infusion of NBQX had begun. All compounds were dissolved in artificial cerebrospinal fluid (ACSF) consisting of (in mM): 124 NaCl, 3.0 KCl, 1 NaH₂PO₄, 1.3 H₂O, 3.0 MgCl₂, 10 dextrose, 10.0 HEPES (pH 7.35), 3.0 CaCl₂. Control experiments involved infusion of ACSF alone.

Histology

The infusion site was marked by passing a direct current anodal current (200 µA for approximately 10 s) through a small wire cut to the length of the internal cannula and exposed at the tip. Animals were killed with an overdose of sodium pentobarbital and

perfused intracardially with 1.0 l of 10% formalin. Brains were embedded in an albumin gelatin mixture, and the brainstem was sectioned using a freezing microtome (80 µm sections). Tissue was mounted, stained with cresyl violet, and counterstained with Prussian blue.

Computer simulations

The circuitry of the simulation was based on the known anatomy and physiology of the cerebellar-olivary system^{6,13} and the way in which this circuitry is engaged during eyelid conditioning^{7,8,15}. Details about the simulation and the physiology of its connections have been presented previously^{14–16}, and will be described only briefly. For each simulated neuron, we used a single-compartment, integrate-and-fire representation, which calculates membrane potential based on leak and synaptic conductances. Of particular relevance to the present results is the connectivity of the inferior olive shown in Fig. 3a. In short, climbing fibres (shown in red) receive inhibitory input from the deep cerebellar nuclei (shown in green) and excitatory input from the reinforcing unconditioned stimulus (shown in black)⁶. Thus, presentation of the unconditioned stimulus during paired trials was simulated by adding a constant depolarizing pulse to the membrane potential of the climbing fibre. In our simulations, acquisition/extinction produce increases/decreases in cerebellar nuclei activity through the induction of plasticity in both the cerebellar cortex and cerebellar nuclei^{14–16}. However, the data presented here regarding the activity of simulated climbing fibres during conditioning only require that cerebellar nuclei activity increases during acquisition (and decreases during extinction), and therefore our results do not depend on the plasticity mechanisms underlying the changes in cerebellar nuclei activity that are observed during eyelid conditioning.

Received 2 July 2001; accepted 2 January 2002.

- Marr, D. A theory of cerebellar cortex. *J. Physiol. (Lond.)* **202**, 437–470 (1969).
- Albus, J. S. A theory of cerebellar function. *Math. Biosci.* **10**, 25–61 (1971).
- Ito, M. Cerebellar control of the vestibulo-ocular reflex—around the flocculus hypothesis. *Annu. Rev. Neurosci.* **5**, 275–296 (1982).
- Thach, W. T., Goodkin, H. P. & Keating, J. G. The cerebellum and the adaptive coordination of movement. *Annu. Rev. Neurosci.* **15**, 403–442 (1992).
- Simpson, J. L., Wylie, D. R. & de Zeeuw, C. I. On climbing fiber signals and their consequence(s). *Behav. Brain Sci.* **19**, 384–398 (1996).
- de Zeeuw, C. I. et al. Microcircuitry and function of the inferior olive. *Trends Neurosci.* **21**, 391–400 (1998).
- Thompson, R. F., Thompson, J. K., Kim, J. J., Krupa, D. J. & Shinkman, P. G. The nature of reinforcement in cerebellar learning. *Neurobiol. Learn. Mem.* **70**, 150–176 (1998).
- Raymond, J. L., Lisberger, S. G. & Mauk, M. D. The cerebellum: a neuronal learning machine? *Science* **272**, 1126–1131 (1996).
- Hesslow, G. & Ivarsson, M. Inhibition of the inferior olive during conditioned responses in the decerebrate ferret. *Exp. Brain Res.* **110**, 36–46 (1996).
- Sears, L. L. & Steinmetz, J. E. Dorsal accessory inferior olive activity diminishes during acquisition of the rabbit classically conditioned eyelid response. *Brain Res.* **545**, 114–122 (1991).
- Kim, J. J., Krupa, D. J. & Thompson, R. F. Inhibitory cerebello-olivary projections and blocking effect in classical conditioning. *Science* **279**, 570–573 (1998).
- Mauk, M. D., Steinmetz, J. E. & Thompson, R. F. Classical conditioning using stimulation of the inferior olive as the unconditioned stimulus. *Proc. Natl Acad. Sci. USA* **83**, 5349–5353 (1986).
- Voogd, J. & Glickstein, M. The anatomy of the cerebellum. *Trends Neurosci.* **21**, 370–375 (1998).
- Medina, J. F., Garcia, K. S., Nore, W. L., Taylor, N. M. & Mauk, M. D. Timing mechanisms in the cerebellum: testing predictions of a large-scale computer simulation. *J. Neurosci.* **20**, 5516–5525 (2000).
- Medina, J. F. & Mauk, M. D. Computer simulation of cerebellar information processing. *Nature Neurosci.* **3** (Suppl.), 1205–1211 (2000).
- Medina, J. F., Garcia, K. S. & Mauk, M. D. A mechanism for savings in the cerebellum. *J. Neurosci.* **21**, 4081–4089 (2001).
- Lang, E. J., Sugihara, I. & Llinas, R. GABAergic modulation of complex spike activity by the cerebellar nucleolus pathway in rat. *J. Neurophysiol.* **76**, 255–275 (1996).
- Demer, J. L., Echelman, D. A. & Robinson, D. A. Effects of electrical stimulation and reversible lesions of the olivocerebellar pathway on Purkinje cell activity in the flocculus of the cat. *Brain Res.* **346**, 22–31 (1985).
- Keating, J. G. & Thach, W. T. Nonclock behavior of inferior olive neurons: interspike interval of Purkinje cell complex spike discharge in the awake behaving monkey is random. *J. Neurophysiol.* **73**, 1329–1340 (1995).
- Ramnanan, N. & Yeo, C. H. Reversible inactivations of the cerebellum prevent the extinction of conditioned initiating membrane responses in rabbits. *J. Physiol. (Lond.)* **495**, 159–168 (1996).
- Linden, D. J. & Connor, J. A. Cellular mechanisms of long-term depression in the cerebellum. *Curr. Opin. Neurobiol.* **3**, 401–406 (1993).
- Sakurai, M. Synaptic modification of parallel fibre-Purkinje cell transmission in *in vitro* guinea-pig cerebellar slices. *J. Physiol. (Lond.)* **394**, 463–480 (1987).
- Hollerman, J. R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neurosci.* **1**, 304–309 (1998).
- Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* **23**, 473–500 (2000).
- Fanselow, M. S. Pavlovian conditioning, negative feedback, and blocking: mechanisms that regulate association formation. *Neuron* **20**, 625–627 (1998).

Acknowledgements

We thank N. Taylor for technical assistance and J. Chin, N. Waxham and J. Knierim for comments on the manuscript.

Correspondence and requests for materials should be addressed to M.M. (e-mail: M.Mauk@uth.tmc.edu).

Oestrogen protects FKBP12.6 null mice from cardiac hypertrophy

Hong-Bo Xin^{*,†}, Takaaki Senbonmatsu[†], Dong-Sheng Cheng^{*}, Yong-Xiao Wang[‡], Julio A. Copello^{*}, Guang-Ju Ji[‡], Mei Lin Collier[§], Ke-Yu Deng[‡], Loice H. Jeyakumar^{*}, Mark A. Magnuson[¶], Tadashi Inagami[†], Michael I. Kotlikoff[‡] & Sidney Fleischer^{*}||

^{*} Department of Biological Sciences, [†] Department of Biochemistry, [¶] Department of Molecular Physiology and Biophysics, || Department of Pharmacology, Vanderbilt University, Nashville, Tennessee 37235, USA

[‡] Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA

[§] Department of Animal Biology, School of Veterinary Medicine, University of Pennsylvania, Pennsylvania 19104, USA

FK506 binding proteins 12 and 12.6 (FKBP12 and FKBP12.6) are intracellular receptors for the immunosuppressant drug FK506 (ref. 1). The skeletal muscle ryanodine receptor (RyR1) is isolated as a hetero-oligomer with FKBP12 (ref. 2), whereas the cardiac ryanodine receptor (RyR2) more selectively associates with FKBP12.6 (refs 3, 4, 5). FKBP12 modulates Ca^{2+} release from the sarcoplasmic reticulum in skeletal muscle^{6,7} and developmental cardiac defects have been reported in FKBP12-deficient mice⁸, but the role of FKBP12.6 in cardiac excitation–contraction coupling remains unclear. Here we show that disruption of the *FKBP12.6* gene in mice results in cardiac hypertrophy in male mice, but not in females. Female hearts are normal, despite the fact that male and female knockout mice display similar dysregulation of Ca^{2+} release, seen as increases in the amplitude and duration of Ca^{2+} sparks and calcium-induced calcium release gain. Female *FKBP12.6*-null mice treated with tamoxifen, an oestrogen receptor antagonist, develop cardiac hypertrophy similar to that of male mice. We conclude that FKBP12.6 modulates cardiac excitation–contraction coupling and that oestrogen plays a protective role in the hypertrophic response of the heart to Ca^{2+} dysregulation.

FKBP12 and FKBP12.6 are tightly associated with skeletal and cardiac ryanodine receptors (RyR1 and RyR2, respectively)^{2,4}, intracellular Ca^{2+} release channels that mediate Ca^{2+} release from intracellular stores in the sarcoplasmic reticulum (SR) and play essential roles in excitation–contraction (E–C) coupling in cardiac and skeletal muscle^{9,10}. Although FKBP12 plays a role in E–C coupling in skeletal muscle by modulating the release of Ca^{2+} from SR^{6,7,11–12}, the effects of FKBP12.6 on cardiac E–C coupling remain uncertain^{4,12}. To determine the functional role of FKBP12.6 in cardiac E–C coupling, the *FKBP12.6* gene was disrupted in embryonic stem cells by homologous recombination using a targeting vector in which exon 3, coding for the binding sites for RyR2, FK506 and calcineurin^{5,13,14}, was replaced by a neomycin-resistance (*neo*^r) cassette (Fig. 1a). The genotype of homozygous *FKBP12.6*^{−/−} mice was confirmed by Southern blot analysis (Fig. 1b) and polymerase chain reaction (PCR) (Fig. 1c); PCR after reverse transcription of RNA (RT–PCR) indicated a lack of FKBP12.6 messenger RNA expression in hearts from knockout mice (Fig. 1d) and that the expression of FKBP12 mRNA is much greater than FKBP12.6 mRNA in wild-type hearts. *FKBP12.6*^{−/−} mice grow at normal rates and both male and female mutants are fertile.

To investigate the cellular consequences of *FKBP12.6* gene inactivation, we examined calcium-induced calcium release (CICR) over a range of depolarization using simultaneous measurements of calcium current, I_{Ca} , and spatially averaged concentration of intracellular calcium ions, $[\text{Ca}^{2+}]_i$, in voltage-clamped, Cs^+ and fura-2 loaded cardiac myocytes from age- and strain-matched male and female mice¹⁵. In some experiments, the length change of myocytes was measured simultaneously using video edge-detection.

Neither the magnitude nor the kinetics of I_{Ca} was altered in FKBP12.6-deficient heart cells relative to wild-type mice, but a marked increase in CICR gain was observed. That is, an equivalent I_{Ca} produced a much larger increase in the peak $[\text{Ca}^{2+}]_i$ transient ($\Delta[\text{Ca}^{2+}]_i$) in FKBP12.6-deficient cardiac myocytes at all voltages examined, indicating a substantially larger release of Ca^{2+} from the SR (Fig. 2). For example, for voltage clamp steps to 10 mV (near the peak of the current–voltage, I – V , curve for I_{Ca}), the $\Delta[\text{Ca}^{2+}]_i$ per picoampere (pA) of I_{Ca} in *FKBP12.6*^{−/−} cardiac myocytes was 83% higher than in wild-type cardiomyocytes from male mice (8.4 nM pA^{−1} versus 4.6 nM pA^{−1}; $P < 0.05$) and similar results were observed in cells from female animals (67% increase over wild type; $P < 0.05$). The increase in CICR gain was not significantly different between male and female mice. Augmented Ca^{2+} release associated with the increase in CICR gain also produced a greater contraction in individual ventricular myocytes. In experiments, as above, in which contraction was also simultaneously measured by edge-detection, depolarization to 10 mV produced a significantly greater shortening in FKBP12.6-deficient myocytes ($21.0 \pm 0.7\%$ versus $10.2 \pm 3.0\%$ for the wild type) ($P < 0.01$, $n = 6$), providing further evidence for an increase in the net cytosolic Ca^{2+} load during E–C coupling.

The molecular basis of the increase in CICR gain was examined by measurements of unitary Ca^{2+} release events in FKBP12.6-deficient and wild-type mice¹⁶. In a series of measurements of individual sparks produced by ramp depolarizations, the amplitude and kinetics of Ca^{2+} sparks were significantly altered in both male and female *FKBP12.6*^{−/−} mice. Ca^{2+} sparks were increased in amplitude (increase in F/F_0 at peak) and size (increase in size of spark at half maximal amplitude, FWHM), and were markedly longer in duration; these sparks were mainly associated with a dramatic increase in decay time in both male (Fig. 3b) and female (Fig. 3c) null mice compared with wild-type mice. The rise time and time constant for

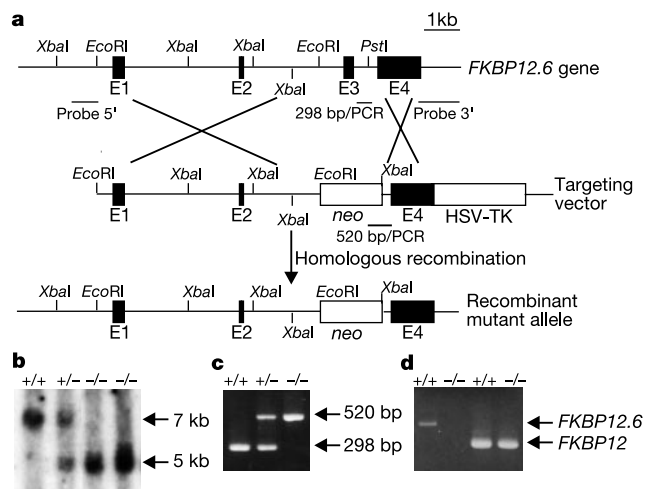


Figure 1 Targeted disruption of *FKBP12.6* gene in mice. **a**, Restriction maps of the mouse *FKBP12.6* gene (top), the targeting vector (middle) and the homologous recombinant mutant allele (bottom). A 1.4-kilobase (kb) *EcoRI*–*PstI* segment that includes exon 3 was replaced with a neomycin-resistance (*neo*^r) expression cassette; a HSV-TK cassette linked to the 3' end and gancyclovir selection was used to eliminate randomly integrated clones. **b**, Southern blot analysis of tail DNA from littermates; the DNA was digested with *XbaI* and hybridized with a 3' probe. The wild-type allele and the mutant mice generate a 7.0-kb and 5.0-kb band, respectively. **c**, Allele-specific PCR analysis. A 298-base pair (bp) and a 520-bp product were generated from the wild-type mice and mutant mice, respectively. **d**, PCR after reverse transcription of RNA (RT–PCR) analysis. A 685-bp fragment was amplified using a pair of mouse *FKBP12.6*-specific primers (from exon 1 and exon 4, respectively) from wild-type (lane 1) but not *FKBP12.6*-null (lane 2) mice. *FKBP12*-specific primers amplified a 483-bp product in wild-type (lane 3) and *FKBP12.6*^{−/−} (lane 4) mice. A threefold greater amount of total RNA was used as a template for FKBP12.6 (lanes 1, 2) relative to FKBP12 RT–PCR (lanes 3, 4).

the rapid decay phase of the Ca^{2+} spark increased significantly in cardiac myocytes from *FKBP12.6*^{-/-} mice compared with wild-type mice. Our results are consistent with the RyR remaining open for longer periods of time (~30% longer) in mice lacking FKBP12.6. Although studies in spontaneously hypertensive rates have indicated that cardiac hypertrophy itself can result in an increase in Ca^{2+} spark amplitude and augmented Ca^{2+} spark size¹⁷, alterations in Ca^{2+} spark kinetics are not a feature of hypertrophy itself. Rather, our results indicate a role for FKBP12.6 in RyR2 gating, resulting in longer channel openings, augmentation of Ca^{2+} release, and the observed increase in CICR gain (Fig. 2).

Despite equivalent alterations in CICR and Ca^{2+} sparks in male and female heart cells, echocardiographic examination¹⁸ of adult (four-month-old) *FKBP12.6*^{-/-} mice indicated a significant increase in cardiac mass relative to age- and strain-matched controls in male but not in female mice (Fig. 4a). The left ventricular mass/body weight ratio (LVM/BW) in male *FKBP12.6*^{-/-} mice was increased by 32% and the thickness of the interventricular septum (IVS) and left ventricular posterior wall (LVPW) were increased by 44% and 22%, respectively, whereas the left ventricular end-diastolic diameter (LVDD) was similar in wild-type and knockout mice. Similar results were obtained in a second study using F₁ mice

derived from a colony maintained as heterozygotes or F₂ derived from F₁ homozygotes. Surprisingly, the female mouse hearts were of normal size, when compared to wild-type mice. These results were confirmed in post-mortem measurements of the hearts of adult mice (Fig. 4b). *FKBP12.6*^{-/-} male mice had higher heart weights (30%) and heart weight/body weight ratios (29%), whereas the heart weights in female knockout mice was not different from those of female wild-type mice. Blood-pressure measurements of male mice at 16–18 weeks of age indicated hypertension in male *FKBP12.6*-null mice (systolic: 145.1 ± 5.0 ($n = 6$) versus 124.7 ± 5.4 ($n = 4$) mmHg, $P < 0.05$; diastolic: 111.3 ± 2.7 ($n = 6$) versus 92.6 ± 4.1 ($n = 4$) mmHg, $P < 0.01$), whereas female null mice were not hypertensive ($n = 5$). Comparison of heart gross morphology revealed larger hearts in male knockout mice when compared to the wild-type (Supplementary Information), as suggested from the larger heart weights (Fig. 4b). Histological evaluation of

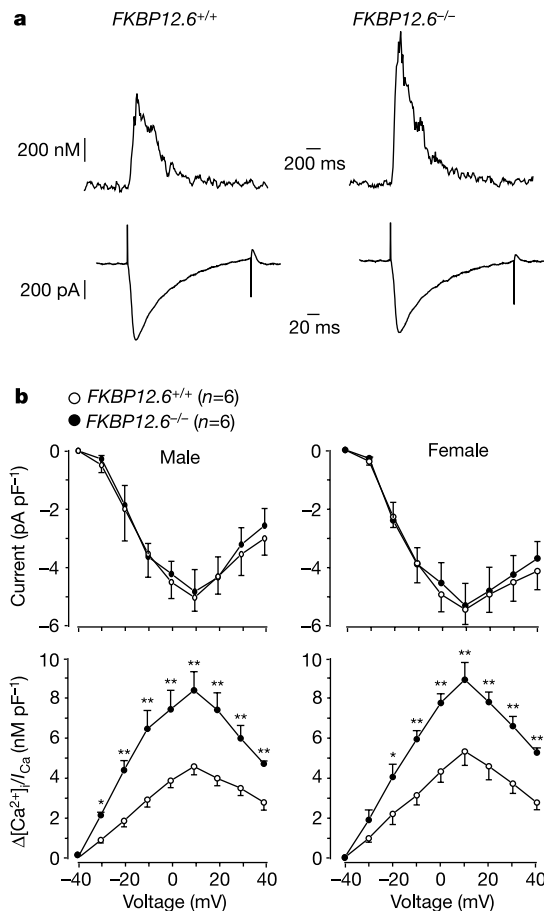


Figure 2 Increased CICR gain in *FKBP12.6* knockout myocytes. **a**, Simultaneous recording of calcium current I_{Ca} (below) and spatially averaged $[\text{Ca}^{2+}]_i$ (above) in a ventricular myocyte depolarized to 10 mV. We note that a similar-magnitude calcium current evokes a substantially greater increase in $[\text{Ca}^{2+}]_i$ in the *FKBP12.6* knockout cell. **b**, Above, I_{Ca} as a function of depolarization voltage in ventricular myocytes isolated from male and female wild-type and knockout mice. Data are normalized by the capacitance of the cell (proportional to cell surface area). Below, peak $[\text{Ca}^{2+}]_i$ achieved following depolarization. The data are normalized by I_{Ca} , indicating that at equivalent I_{Ca} , male and female knockout cells release more Ca^{2+} than the wild type. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$.

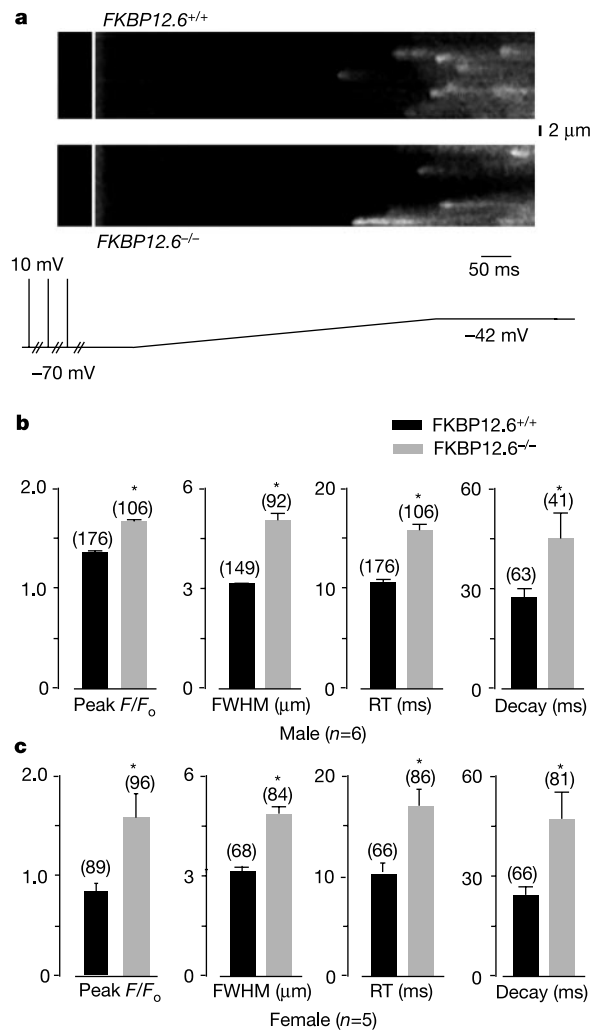


Figure 3 Altered Ca^{2+} sparks in *FKBP12.6* knockout mice. **a**, Line-scan mode recording from voltage-clamped wild-type and knockout myocytes. Pre-stimulation and voltage ramp protocol from -70 to -42 mV are shown below. A light-emitting diode was flashed at the initiation of the voltage ramp (white line on line-scan traces above). Calcium sparks in knockout cells often displayed long tails indicating delayed inactivation (see first spark in *FKBP*^{-/-}). Summary of characteristics of individual calcium sparks from male (**b**) and female (**c**) wild-type versus knockout cells indicate significant difference in amplitude and kinetics of calcium sparks in *FKBP12.6*-null cardiomyocytes. F/F_0 , fluorescence magnitude divided by average pre-stimulus fluorescence (unit-less); FWHM, full width at half maximum spark size (μm); RT, rise time (taken from point of rise to peak F/F_0); Decay, time constant in ms (exponential fit of rapid phase of F/F_0 decay); number of sparks analysed indicated in parentheses (asterisk indicates $P < 0.05$).

four-month-old mice did not reveal connective tissue infiltration; however, we did observe occasional focal disarray in male, but not female, *FKBP12.6*^{-/-} mice in the upper region of the interventricular septum (Supplementary Information), which was not observed in wild-type mice. Such focal disarray has previously been reported in the interventricular septum of humans with hypertrophic cardiomyopathy¹⁹.

Because equivalent dysregulation of Ca²⁺ release was observed in male and female *FKBP12.6*^{-/-} mice, but only male mice developed enlarged hearts, we reasoned that one or more factors protected female myocytes from the hypertrophic effects of Ca²⁺ overload. Women experience a marked increase in the incidence of left ventricular hypertrophy after menopause²⁰ and hormonal replacement therapy in postmenopausal women results in a significantly reduced left ventricular mass relative to age-matched controls²¹. To determine whether oestrogens play a protective role in the cellular response to Ca²⁺ dysregulation, we examined heart size parameters in adult female mice continually exposed to the oestrogen receptor antagonist tamoxifen. As shown in Fig. 4c, female

FKBP12.6-null mice treated with tamoxifen developed cardiac hypertrophy relative to wild-type (treated and untreated) and untreated knockout mice. Moreover, the degree of cardiac hypertrophy in female *FKBP12.6*-null mice treated with tamoxifen was quite similar to hypertrophy in male mice. For example, in knockout females treated with tamoxifen, the LVM/BW ratio was 35% greater than the knockout controls and 27% greater than the wild-type controls, whereas tamoxifen-treated wild-type animals were not significantly different from untreated wild-type or knockout mice. Similar significant differences were noted for IVS and LVPW. Tamoxifen treatment itself did not affect any cardiac parameters in the wild-type female mice, suggesting a specific effect in Ca²⁺ dysregulated cardiac myocytes. Thus oestrogens play a protective role in the hypertrophic response of heart cells to *FKBP12.6* deletion.

Our studies establish an important role for the *FKBP12.6* gene product in cardiac E-C coupling. Deletion of the *FKBP12.6* gene produces an increase in Ca²⁺ spark duration and an increase in CICR gain, resulting in an increased Ca²⁺ load in both male and female cardiac ventricular myocytes. A decrease in CICR gain has been demonstrated in failing hearts²², and this system appears to play a central role in transducing mechanical and cellular signals in the failing human heart²³. Despite the similar alteration in CICR gain observed in male and female *FKBP12.6*^{-/-} mice, male hearts undergo hypertrophy and display focal disarray in the IVS, whereas the gross morphology and histology of the hearts from female *FKBP12.6*^{-/-} mice are normal. The simplest explanation of these data is that the adaptive response of the heart to Ca²⁺ overload is regulated by oestrogen receptor signalling. Increases in [Ca²⁺]_i augment calcineurin activity, resulting in the nuclear translocation of NFATc transcription factors and the initiation of a programme of gene expression that leads to cardiac hypertrophy²⁴. The degree to which the expression of these genes is modulated by sex-hormone-dependent processes is not known. We note, however, that tumour-necrosis factor- α (TNF- α) expression is markedly upregulated by NFAT transcription factors²⁵ and overexpression of TNF- α in transgenic mice results in cardiomyopathy in male but not female mice, presumably owing to the greater expression of TNF receptors in male hearts²⁶. Thus augmented CICR gain in *FKBP12.6*^{-/-} mice could trigger gene expression that results in cardiac hypertrophy in male but not in female mice. However, we cannot exclude triggering mechanisms independent of alterations in cardiac CICR, including changes in calcineurin activity due to loss of *FKBP12.6* binding or increases in vascular peripheral resistance in *FKBP12.6*-null mice resulting in systematic hypertension and secondary cardiac hypertrophy. Regardless of the initiating mechanism, our findings may be viewed in the context of human epidemiological data indicating gender-specific effects on cardiac hypertrophy suggestive of a hormone-related effect on the left ventricular mass²⁷. While the relationships between specific sex hormones, gender-specific cardiovascular diseases and the protective value of specific hormone-replacement therapies is unclear²⁸, epidemiological and animal studies suggest a strong link between female hormonal status and cardiac size²⁹. *FKBP12.6* knockout mice, which develop hypertensive cardiac hypertrophy in a sex-specific manner, are likely to facilitate investigation of the signalling programme triggered by Ca²⁺ overload and the processes by which sex hormones modulate these effects. □

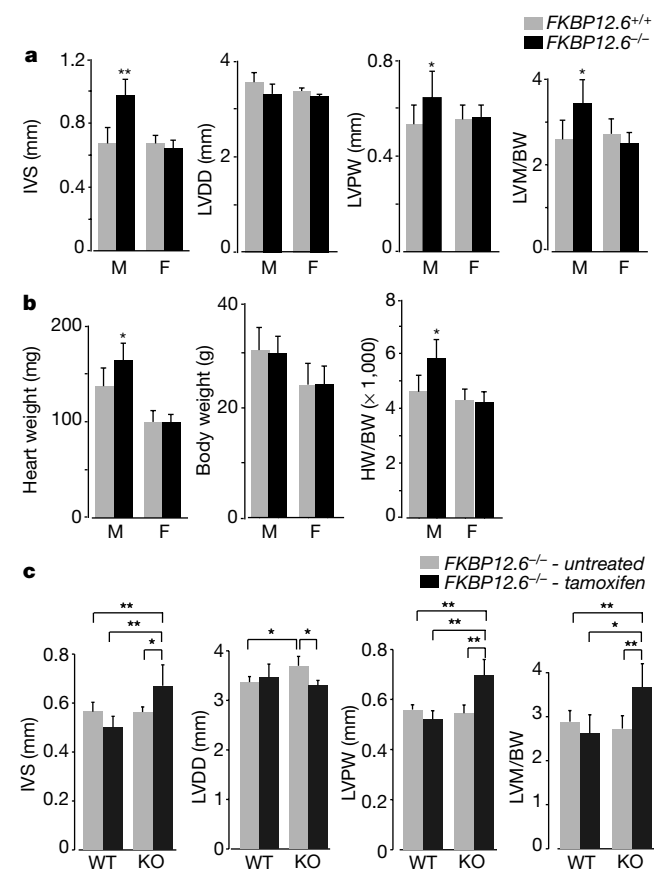


Figure 4 Sex-specific cardiac hypertrophy in male *FKBP12.6* knockout mice and cardiac hypertrophy in female knockout mice treated with tamoxifen. **a**, Echocardiographic measurements in adult *FKBP12.6*^{-/-} mice ($n = 8$ in male (M), $n = 11$ in female (F)) and wild type mice ($n = 8$ in male and female). IVS, interventricular septum; LVDD, left ventricular end-diastolic diameter; LVPW, left ventricular posterior wall; and LVM/BW, left ventricular mass/body weight ratio. **b**, Comparison of heart weights in wild-type mice ($n = 6$ in male and $n = 10$ in female) and *FKBP12.6*^{-/-} ($n = 6$ in male and $n = 30$ in female) mice. HW/BW, heart weight/body weight ratio. Data are presented as mean $\pm$ s.e.; within-sex significance was determined by Student's *t*-test (asterisk, $P < 0.05$; double asterisk, $P < 0.01$). **c**, Cardiac hypertrophy in *FKBP12.6* knockout female mice treated with tamoxifen. Female null mice treated with tamoxifen have significant hypertrophy relative to untreated knockout mice and treated and untreated wild-type mice. Each group contains 6–8 animals; statistical analysis was performed using ANOVA to compare the four groups (null and wild-type, treated and untreated; asterisk, $P < 0.05$; double asterisk, $P < 0.01$). WT, wild type; KO, knockout.

Methods

Generation of *FKBP12.6*-deficient mice

An *FKBP12.6* genomic clone encoding the entire *FKBP12.6* gene was isolated from a murine 129/SvEv genomic DNA library (Stratagene) using mouse *FKBP12.6* complementary DNA as a probe. The targeting vector contained PGK-neo flanked by *FKBP12.6* genomic DNA and an HSV-TK cassette³⁰ and resulted in replacement of exon 3 with the *neo*^r expression cassette in homologous recombinants selected with G418 and gancyclovir. Chimaeric male mice obtained from two selected embryonic stem cell clones were bred to either 129/SvEv female mice to establish an inbred line, or to C57BL/6 female

mice to establish a hybrid line. Homozygous null mice were obtained by heterozygous mating.

Single myocyte methods

Single ventricular myocytes from 4–6-month-old mice were prepared by Langendorff perfusion¹⁶. I_{Ca} and $[Ca^{2+}]_i$ were measured simultaneously in voltage-clamped myocytes. Spatially averaged measurements of $[Ca^{2+}]_i$ were made in myocytes dialysed with fura-2 (75 μ M)¹⁵, and Ca^{2+} spark measurements in myocytes dialysed with fluo-4 (100 μ M) and imaged with a Noran (Oz) or BioRad (Radiance 2000) laser scanning confocal microscope. The Ca^{2+} indicator was added to the intracellular solution (mm): 127 Cs-glutamate, 10 HEPES, 19 TEACl, 0.1 Tris-GTP, 1 Mg-ATP, and 5 Tris₂-creatine phosphate at pH 7.2. The extracellular solution was (mM): 137 NaCl, 5 CsCl, 2 CaCl₂, 1 NaH₂PO₄, 1.2 MgCl₂, 10 HEPES and 5 glucose at pH 7.4. The voltage protocol maximized SR loading (multiple pre-depolarizations) and minimized the Na⁺ current (holding at -40 mV). In some experiments, myocyte length was simultaneously monitored by edge detection.

In vivo measurements

Echocardiography measurements were performed in age-matched adult wild-type and FKBP12.6^{-/-} mice as described¹⁸. Two-dimensional views of the left ventricle were obtained to guide M-mode measurements of the IVS (interventricular septum), LVPW (left ventricular posterior wall), and LVDD (left ventricular end-diastolic diameter). Three measurements were averaged for each parameter in every mouse. The echocardiographic LVM (left ventricular mass) was calculated using the specific gravity of myocardium (1.055 mg mm⁻³) and the cube formula: $LVM = 1.055 \times [(IVS + LVPW + LVDD)^3 - (LVDD)^3]$. Statistical comparisons were made by analysis of variance (ANOVA). Blood pressure was measured in conscious animals with an indwelling cannula in the carotid artery; measurements were repeated five times and averaged for 5 min; readings at 24 and 48 hours agreed well.

Morphological analysis

After euthanizing the mice, their chests were opened and the hearts perfused with phosphate-buffered saline (PBS). Hearts were then removed and fixed overnight in 10% formalin buffered with PBS, dehydrated in ethanol, transferred to xylene, cut transversely into upper, mid and apex regions of the heart and aligned and embedded in paraffin, sectioned transversely, and stained for histological examination.

Administration of tamoxifen to female mice

Tamoxifen slow-release pellets (NE-361, Innovative Research of America) designed to release 2–2.5 ng ml⁻¹ blood levels for a period of 90 days were implanted between the skin and muscle in the back of the neck of mice at puberty (6–7-week-old female wild-type and knockout mice). Echocardiography measurements were performed on these mice after the pellets had been implanted for 10 weeks.

Received 17 September; accepted 17 December 2001.

1. Marks, A. R. Cellular functions of immunophilins. *Physiol. Rev.* **76**, 631–649 (1996).
2. Jayaraman, T. *et al.* FK506 binding protein associated with the calcium release channel (ryanodine receptor). *J. Biol. Chem.* **267**, 9474–9477 (1992).
3. Lam, E. *et al.* A novel FK506 binding protein can mediate the immunosuppressive effects of FK506 and is associated with the cardiac ryanodine receptor. *J. Biol. Chem.* **270**, 26511–26522 (1995).
4. Timmerman, A. P. *et al.* Selective binding of FKBP12.6 by the cardiac ryanodine receptor. *J. Biol. Chem.* **271**, 20385–20391 (1996).
5. Xin, H. B., Rogers, K., Qi, Y., Kanematsu, T. & Fleischer, S. Three amino acid residues determine selective binding of FK506-binding protein 12.6 to the cardiac ryanodine receptor. *J. Biol. Chem.* **274**, 15315–15319 (1999).
6. Timmerman, A. P. *et al.* The calcium release channel of sarcoplasmic reticulum is modulated by FK-506-binding protein. Dissociation and reconstitution of FKBP-12 to the calcium release channel of skeletal muscle sarcoplasmic reticulum. *J. Biol. Chem.* **268**, 22992–22999 (1993).
7. Brillantes, A. B. *et al.* Stabilization of calcium release channel (ryanodine receptor) function by FK506-binding protein. *Cell* **77**, 513–523 (1994).
8. Shou, W. *et al.* Cardiac defects and altered ryanodine receptor function in mice lacking FKBP12. *Nature* **391**, 489–492 (1998).
9. Fleischer, S. & Inui, M. Biochemistry and biophysics of excitation–contraction coupling. *Annu. Rev. Biophys. Chem.* **18**, 333–364 (1989).
10. Meissner, G. Ryanodine receptor/Ca²⁺ release channels and their regulation by endogenous effectors. *Annu. Rev. Physiol.* **56**, 485–508 (1994).
11. Timmerman, A. P., Wiederrecht, G., Marcy, A. & Fleischer, S. Characterization of an exchange reaction between soluble FKBP-12 and the FKBP-ryanodine receptor complex. Modulation by FKBP mutants deficient in peptidyl-prolyl isomerase activity. *J. Biol. Chem.* **270**, 2451–2459 (1995).
12. Barg, S., Copello, J. A. & Fleischer, S. Different interactions of cardiac and skeletal muscle ryanodine receptors with FK-506 binding protein isoforms. *Am. J. Physiol.* **272**, C1726–C1733 (1997).
13. Van Duyn, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. *Science* **252**, 839–842 (1991).
14. Aldape, R. A. *et al.* Charged surface residues of FKBP12 participate in formation of the FKBP12-FK506-calcineurin complex. *J. Biol. Chem.* **267**, 16029–16032 (1992).
15. Wang, Y.-X. & Kotlikoff, M. I. Inactivation of calcium-activated chloride channels in smooth muscle by calcium/calmodulin dependent protein kinase. *Proc. Natl Acad. Sci. USA* **94**, 14918–14923 (1997).
16. Collier, M. L., Ji, G., Wang, Y. & Kotlikoff, M. I. Calcium-induced calcium release in smooth muscle: loose coupling between the action potential and calcium release. *J. Gen. Physiol.* **115**, 653–662 (2000).
17. Shorofsky, S. R. *et al.* Cellular mechanisms of altered contractility in the hypertrophied heart: big hearts, big sparks. *Circ. Res.* **84**, 424–434 (1999).
18. Fentzke, R. C. *et al.* Evaluation of ventricular and arterial hemodynamics in anesthetized closed-chest mice. *J. Am. Soc. Echocardiogr.* **10**, 915–925 (1997).
19. Louie, E. K. & Edwards, L. C. III Hypertrophic cardiomyopathy. *Prog. Cardiovasc. Dis.* **36**, 275–308 (1994).

20. Levy, D. *et al.* Echocardiographically detected left ventricular hypertrophy: prevalence and risk factors. The Framingham Heart Study. *Ann. Intern. Med.* **108**, 7–13 (1988).
21. Lim, W. K., Wren, B., Jenson, N., Roy, S. & Caplan, G. Effect of hormone replacement therapy on left ventricular hypertrophy. *Am. J. Cardiol.* **83**, 1132–1134 (1999).
22. Gomez, A. M. *et al.* Defective excitation–contraction coupling in experimental cardiac hypertrophy and heart failure. *Science* **276**, 800–806 (1997).
23. Marx, S. O. *et al.* PKA phosphorylation dissociates FKBP12.6 from the calcium release channel (ryanodine receptor): defective regulation in failing hearts. *Cell* **101**, 365–376 (2000).
24. Molkenin, J. D. *et al.* A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* **93**, 215–228 (1998).
25. Falvo, J. V. *et al.* Stimulus-specific assembly of enhancer complexes on the tumor necrosis factor alpha gene promoter. *Mol. Cell Biol.* **20**, 2239–2247 (2000).
26. Kadokami, T., McTiernan, C. E., Kubota, T., Frye, C. S. & Feldman, A. M. Sex-related survival differences in murine cardiomyopathy are associated with differences in TNF-receptor expression. *J. Clin. Invest.* **106**, 589–597 (2000).
27. Pelzer, T., Shamim, A., Wolfes, S., Schumann, M. & Neyes, L. Modulation of cardiac hypertrophy by estrogens. *Adv. Exp. Med. Biol.* **432**, 83–89 (1997).
28. Hulley, S. *et al.* Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *J. Am. Med. Assoc.* **280**, 605–613 (1998).
29. Hayward, C. S., Kelly, R. P. & Collins, P. The roles of gender, the menopause and hormone replacement on cardiovascular function. *Cardiovasc. Res.* **46**, 28–49 (2000).
30. Tybulewicz, V. L., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene. *Cell* **65**, 1153–1163 (1991).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

Histology was carried out at the Vascular Histology and Imaging Core Vanderbilt Division of Cardiology. We thank J. V. Barnett and L. Gleaves for advice and help in preparing the histological sections, J. B. Atkinson for analysing the histological sections, and E. Price for blood-pressure measurements. The production of the FKBP12.6 knockout mouse was achieved with the aid of the Vanderbilt Transgenic/ES Cell Shared Resource Lab. We thank M. Sutherland for help in this regard. This work was supported by grants from the NIH to S.F., T.I., M.M., Y.-X.W. and M.I.K., and a Discovery Grant from Vanderbilt University (S.F.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.F. (e-mail: sidney.fleischer@vanderbilt.edu) or M.I.K. (e-mail: mik7@cornell.edu).

Nitric oxide regulates the heart by spatial confinement of nitric oxide synthase isoforms

Lili A. Barouch*†, Robert W. Harrison*†, Michel W. Skaf*†, Gisele O. Rosas*, Thomas P. Cappola*, Zouflicar A. Kobeissi*, Ion A. Hobai*, Christopher A. Lemmon‡, Arthur L. Burnett§, Brian O'Rourke*, E. Rene Rodriguez||, Paul L. Huang¶, João A. C. Lima*, Dan E. Berkowitz‡# & Joshua M. Hare*

* Departments of Medicine (Cardiology Division), ‡ Biomedical Engineering, § Urology, || Pathology, and # Anaesthesiology and Critical Care Medicine, The Johns Hopkins Medical Institutions, Baltimore, Maryland 21287, USA ¶ Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA

† These authors contributed equally to this work

Subcellular localization of nitric oxide (NO) synthases with effector molecules is an important regulatory mechanism for NO signalling¹. In the heart, NO inhibits L-type Ca²⁺ channels² but stimulates sarcoplasmic reticulum (SR) Ca²⁺ release^{3–5}, leading to variable effects on myocardial contractility. Here we show that spatial confinement of specific NO synthase isoforms

regulates this process. Endothelial NO synthase (NOS3) localizes to caveolae^{6–8}, where compartmentalization with β -adrenergic receptors and L-type Ca^{2+} channels⁹ allows NO to inhibit β -adrenergic-induced inotropy^{8,10}. Neuronal NO synthase (NOS1), however, is targeted to cardiac SR¹¹. NO stimulation of SR Ca^{2+} release via the ryanodine receptor (RyR) *in vitro*^{3,4} suggests that NOS1 has an opposite, facilitative effect on contractility. We demonstrate that NOS1-deficient mice have suppressed inotropic response, whereas NOS3-deficient mice have enhanced contractility, owing to corresponding changes in SR Ca^{2+} release. Both $\text{NOS1}^{-/-}$ and $\text{NOS3}^{-/-}$ mice develop age-related hypertrophy, although only $\text{NOS3}^{-/-}$ mice are hypertensive. $\text{NOS1/3}^{-/-}$ double knockout mice have suppressed β -adrenergic responses and an additive phenotype of marked ventricular remodelling. Thus, NOS1 and NOS3 mediate independent, and in some cases opposite, effects on cardiac structure and function.

To demonstrate specific subcellular localization of NO synthases in cardiac myocardium, we assessed protein–protein interactions between each NO synthase isoform and either RyR or caveolin-3 (Cav3). Total heart protein extracts immunoprecipitated with anti-RyR exhibited immunoreactivity with NOS1, but not NOS3. Conversely, protein extracts immunoprecipitated with Cav3 selectively

bound to NOS3, but not NOS1 (see Supplementary Information Fig. 1). We confirmed unchanged NOS3 abundance in $\text{NOS1}^{-/-}$ hearts and vice versa by immunoblotting (data not shown). These data demonstrate precise spatial confinement of NOS1 and NOS3 in cardiomyocytes, with selective association of NOS1 with its SR effector protein, RyR, and NOS3 with the myocyte caveolae scaffolding protein, Cav3.

Next, to show the independent and opposite roles of NOS1 and NOS3 in regulating cardiac function, we measured cardiovascular performance in mice with homozygous deletions of NOS1 ($\text{NOS1}^{-/-}$), NOS3 ($\text{NOS3}^{-/-}$), both isoforms ($\text{NOS1/3}^{-/-}$), and corresponding wild-type (C57Bl/6) controls (2–5 months old) using a combined left ventricular (LV) pressure–volume catheter¹² (Table 1, Fig. 1a; see also Supplementary Information Table 1). $\text{NOS1}^{-/-}$ hearts ($n = 10$) had similar systolic blood pressure, chamber sizes, and basal contractility compared to C57Bl/6 hearts ($n = 20$). Although $\text{NOS1}^{-/-}$ had mild elevation of peak rate of rise in LV pressure normalized to instantaneous developed pressure, $(dP/dt)/P_{\text{id}}$ (an isovolumic contractility index), coupling of ventricular elastance (E_{es}) to arterial elastance (E_{a})¹³ was unchanged from C57Bl/6 (ratio of $E_{\text{es}}/E_{\text{a}} \approx 1$), indicating normal baseline ventricular performance. In contrast, $\text{NOS3}^{-/-}$ mice ($n = 24$) had higher systolic blood pressure and E_{a} (consistent with the role of NOS3 in maintaining vascular tone), mildly elevated heart rates, and smaller LV chambers. The $E_{\text{es}}/E_{\text{a}}$ ratio remained around 1, indicating normal coupling of the LV to the vasculature. $\text{NOS1/3}^{-/-}$ mice ($n = 17$) had identical blood pressure elevation compared to $\text{NOS3}^{-/-}$, but had elevated heart rate and basal contractility (higher $(dP/dt)/P_{\text{id}}$ and $E_{\text{es}}/E_{\text{a}}$ ratio of ~ 3).

We next determined the β -adrenergic-stimulated cardiac reserve. Isoproterenol infusion ($5\text{--}100 \text{ ng kg}^{-1} \text{ min}^{-1}$) increased myocardial contractility, both by E_{es} (Fig. 1a and b) and $(dP/dt)/P_{\text{id}}$ (Fig. 1c)¹³. Compared to C57Bl/6, $\text{NOS1}^{-/-}$ mice had suppressed β -adrenergic inotropic responses, indicating that NOS1 facilitates myocardial contractile reserve. As previously reported, $\text{NOS3}^{-/-}$ animals had augmented responses to isoproterenol^{12,14}, consistent with NOS3 inhibition of β -adrenergic inotropic responses. We then examined whether the $\text{NOS1}^{-/-}$ phenotype was apparent in the absence of NOS3. NOS1 is downstream of NOS3 in the process of calcium-induced calcium release, which drives excitation-contraction coupling, so we anticipated that $\text{NOS1/3}^{-/-}$ mice would have a $\text{NOS1}^{-/-}$ phenotype with regard to β -adrenergic inotropic response. Indeed, despite elevated basal contractility, the inotropic response of $(dP/dt)/P_{\text{id}}$ was suppressed in $\text{NOS1/3}^{-/-}$ mice compared to $\text{NOS3}^{-/-}$ mice (Fig. 1c, right). These data demonstrate that NOS1 and NOS3, despite similar substrates, cofactors, and Ca^{2+} -dependence, mediate directionally opposite influences on myocardial contractile reserve.

Quantitative polymerase chain reaction (PCR) showed an increase in $\beta 1$ adrenergic receptor (AR) mRNA concentration in $\text{NOS1}^{-/-}$ and $\text{NOS3}^{-/-}$ mice compared to C57Bl/6, although $\beta 2$ -AR mRNA was unchanged (data not shown). However, these

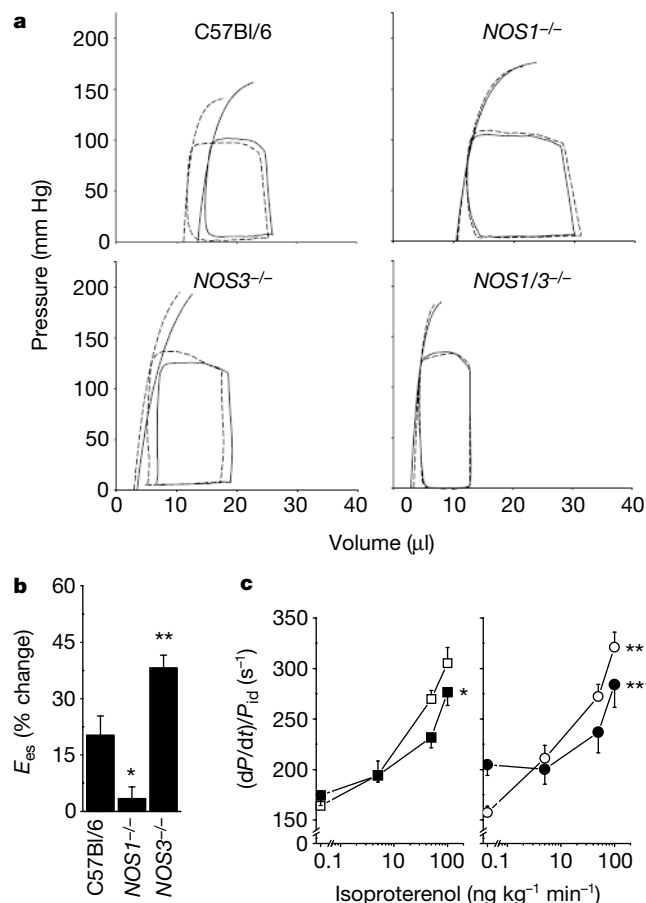


Figure 1 Haemodynamic response to β -adrenergic stimulation. **a**, Pressure–volume loops before (solid curves) and during (dashed curves) isoproterenol ($5 \text{ ng kg}^{-1} \text{ min}^{-1}$). The steeper and left-shifted E_{es} line signifies increased contractility. **b**, E_{es} change with isoproterenol ($5 \text{ ng kg}^{-1} \text{ min}^{-1}$). Cavity obliteration precluded accurate measurements in $\text{NOS1/3}^{-/-}$ mice. (asterisk, $P < 0.05$ versus C57Bl/6; double asterisk, $P < 0.01$ versus C57Bl/6.) **c**, $(dP/dt)/P_{\text{id}}$ isoproterenol response (open squares, C57Bl/6; filled squares, $\text{NOS1}^{-/-}$; open circles, $\text{NOS3}^{-/-}$; filled circles, $\text{NOS1/3}^{-/-}$; $n = 4\text{--}9$ mice for each phenotype). Attenuated $\text{NOS1}^{-/-}$ response versus C57Bl/6 (left; asterisk, $P < 0.01$). Enhanced $\text{NOS3}^{-/-}$ response versus C57Bl/6 (double asterisk, $P < 0.001$) and attenuated $\text{NOS1/3}^{-/-}$ response versus $\text{NOS3}^{-/-}$ (right; triple asterisk, $P < 0.005$).

Table 1 Baseline haemodynamic data

	C57Bl/6	$\text{NOS1}^{-/-}$	$\text{NOS3}^{-/-}$	$\text{NOS1/3}^{-/-}$
<i>n</i>	20	10	24	17
Heart rate (b.p.m.)	573 ± 11	611 ± 9	612 ± 13*	658 ± 11*
Systolic blood pressure (mm Hg)	108 ± 2	113 ± 2	129 ± 4*†	132 ± 4*†
End-diastolic volume (μl)	24.8 ± 2.9	26.5 ± 0.7	12.8 ± 1.3*†	16.4 ± 3.1*†
$(dP/dt)/P_{\text{id}}$ (s ⁻¹)	160.3 ± 3.3	178.4 ± 8.0*	154.5 ± 4.7†	198.3 ± 6.4*††
$E_{\text{es}}/E_{\text{a}}$	0.78 ± 0.08	0.96 ± 0.14	1.36 ± 0.32	3.08 ± 1.23*††
Myocyte size (μm)	8.2 ± 0.2	13.7 ± 0.5*	16.9 ± 0.5*†	11.8 ± 0.4*††

$(dP/dt)/P_{\text{id}}$, peak rate of rise in ventricular pressure normalized to instantaneous developed pressure; $E_{\text{es}}/E_{\text{a}}$, ratio of end-systolic pressure–volume relationship to effective arterial afterload.

* $P < 0.05$ versus C57Bl/6; † $P < 0.05$ versus $\text{NOS1}^{-/-}$; ‡ $P < 0.05$ versus $\text{NOS3}^{-/-}$.

differences did not alter total β -AR density by radioligand binding, which was similar among strains (data not shown). Our findings in the $NOS3^{-/-}$ phenotype agree with those of ref. 14. That mRNA is actually increased and β -AR density is unchanged in $NOS1^{-/-}$ mice confirm that the suppressed responses in $NOS1^{-/-}$ mice are not due to reduction of β -AR expression.

On the basis of the subcellular locale of these two NO synthase isoforms, we hypothesized that the divergent contractile responses were due to altered myocyte Ca^{2+} cycling (Fig. 2a–d). Baseline Ca^{2+} -transients ($[Ca^{2+}]_i$) and sarcomere shortening were similar among strains (data not shown), suggesting that the elevations in basal contractility among intact $NOS1/3^{-/-}$ animals are due to factors extrinsic to the cardiomyocyte (for example, sympathetic tone). Similar to our *in vivo* observations, $NOS1^{-/-}$ myocytes had attenuated responses, while $NOS3^{-/-}$ myocytes had enhanced responsiveness to isoproterenol in $[Ca^{2+}]_i$ and sarcomere shortening (Fig. 2a and c). $NOS1/3^{-/-}$ myocyte responses were decreased by the same proportion relative to $NOS3^{-/-}$ as $NOS1^{-/-}$ compared to C57Bl/6 (Fig. 2b and d; see also Supplementary Information Fig. 2). Thus, contractile responses to β -adrenergic stimulation in isolated myocytes paralleled those of the intact heart. These data demonstrate that the two isoforms have divergent influences on myocyte Ca^{2+} cycling and thus, cardiac contraction. Furthermore,

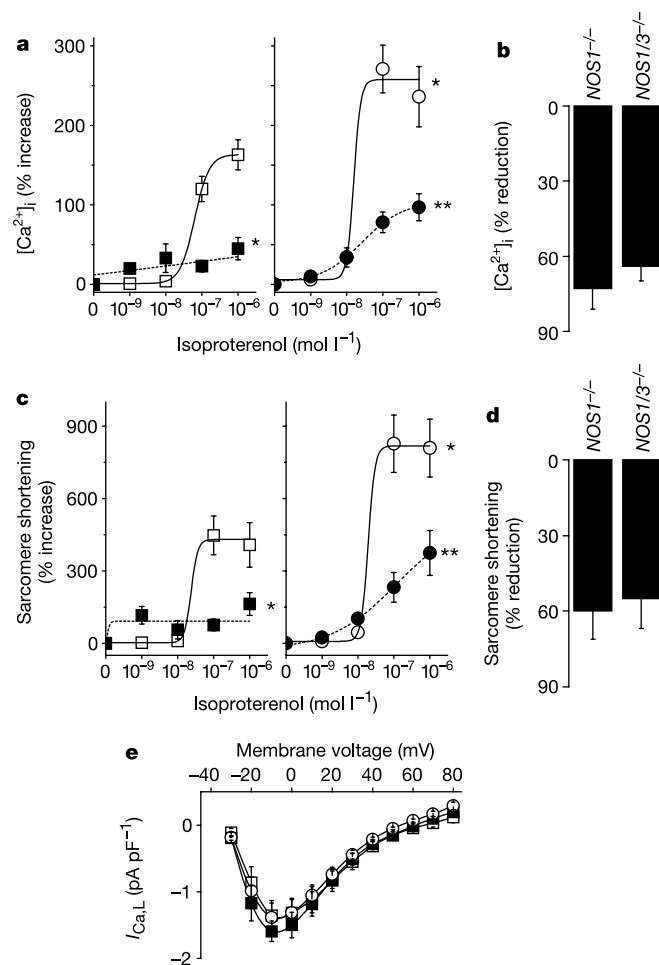


Figure 2 Myocyte response to β -adrenergic stimulation. **a, c**, Isoproterenol concentration-effect response for $[Ca^{2+}]_i$ (**a**) and sarcomere shortening (**c**, asterisk, $P < 0.05$ versus C57Bl/6; double asterisk, $P < 0.005$ versus $NOS3^{-/-}$; $n = 10$ –16 cells, 3–6 hearts each; symbols as in Fig. 1). **b, d**, Reduction in isoproterenol response of $NOS1^{-/-}$ versus C57Bl/6 mice (left) and $NOS1/3^{-/-}$ versus $NOS3^{-/-}$ mice (right) in $[Ca^{2+}]_i$ (**b**) and sarcomere shortening (**d**, P not significant for both). **e**, Similar voltage dependence of I_{CaL} among strains ($n > 8$ cells, 2 hearts each).

the $NOS1^{-/-}$ phenotype is preserved whether or not NOS3 is present.

We also examined the impact of NOS1 or NOS3 deficiency on the β 3-AR signal transduction pathways that activate cardiac NO signalling^{12,15}. BRL37344, a selective β 3-agonist, reduced both $[Ca^{2+}]_i$ and sarcomere shortening in C57Bl/6 myocytes. Responses were unchanged in $NOS1^{-/-}$ myocytes, but absent in those of $NOS3^{-/-}$, supporting linkage of NOS3, but not NOS1, with β -adrenergic signalling mechanisms (see Supplementary Information Fig. 3). Patch-clamp measurements of L-type Ca^{2+} currents (I_{CaL}) at baseline (Fig. 2e) and in response to isoproterenol ($1 \mu M$) in $NOS1^{-/-}$ cells ($47 \pm 1\%$) versus C57Bl/6 ($46 \pm 7\%$; P not significant) were similar. Together these findings further support selective coupling of adrenergic signal transduction and L-type channel activation specifically with NOS3, as predicted by the subcellular localization of this isoform.

As NOS1 and NOS3 have divergent effects on cardiac function, we tested whether they exerted independent influences on cardiac structure. Because disruption of SR Ca^{2+} release can cause cardiac hypertrophy^{16–18}, we predicted that $NOS1^{-/-}$ mice would develop age-related LV hypertrophy, reflected by increased LV wall thickness and mass. $NOS3^{-/-}$ mice also develop LV hypertrophy, although it is associated with hypertension¹⁹. Echocardiography in $NOS1^{-/-}$ ($n = 28$), $NOS3^{-/-}$ ($n = 32$) and C57Bl/6 ($n = 23$) mice demonstrated that although each strain had similar LV mass and LV wall thickness at 2 months, $NOS1^{-/-}$ and $NOS3^{-/-}$ mice developed similar age-related increases in both parameters (Fig. 3a and b; see also Supplementary Information Fig. 4). Importantly, $NOS1^{-/-}$ animals had similar blood pressure to C57Bl/6 (measured under identical anaesthesia), which did not change with age (113 ± 2 versus 99 ± 1 mm Hg in young and old, respectively, P not significant).

To test the interaction of NOS1 and NOS3 in modulating cardiac architecture, we studied $NOS1/3^{-/-}$ mice ($n = 43$). These animals, despite having similar systolic blood pressure to $NOS3^{-/-}$, developed a unique age-related phenotype characterized by concentric remodelling, with marked increase in relative wall thickness (ratio of LV wall thickness to LV radius)²⁰ (Fig. 3c). This was due to increased LV wall thickness and smaller systolic (1.4 ± 0.2 versus 2.3 ± 0.2 mm in C57Bl/6; $P < 0.05$) and diastolic (2.6 ± 0.4 versus 3.6 ± 0.2 mm in C57Bl/6; $P < 0.05$) LV dimensions with age. This structural phenotype was associated with hypercontractility (Table 1) and increased diastolic stiffness, closely resembling hypertensive hypertrophic cardiomyopathy of the elderly in humans²¹ (L.A.B. and J.M.H., manuscript in preparation). Thus, cumulative loss of NOS1 and NOS3 leads to a phenotype not observed in either single

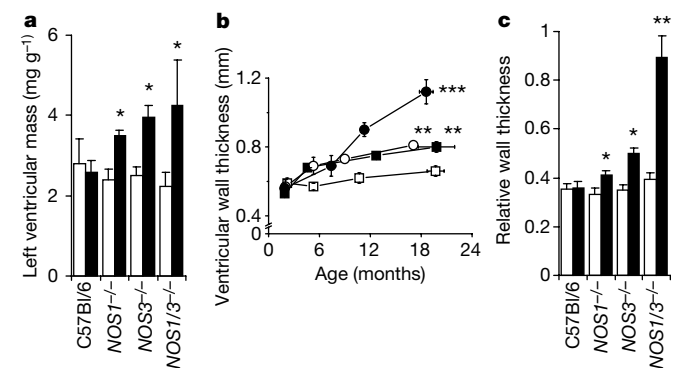


Figure 3 Age-related alterations in left ventricular architecture. **a–c**, Echocardiographic changes in left ventricular mass (**a**), ventricular wall thickness (**b**), and relative wall thickness (**c**), in young (2 months, open bars) and old (14–20 months, black bars) mice (asterisk, $P < 0.05$ versus young and C57Bl/6; double asterisk, $P < 0.0005$ versus C57Bl/6; triple asterisk, $P < 0.0005$ versus $NOS1^{-/-}$ and $NOS3^{-/-}$ by likelihood ratio test; symbols for **b** as in Fig. 1).

NO synthase knockout, supporting independent roles for each isoform in maintaining normal cardiac architecture.

All mutant strains developed myocyte hypertrophy by 5 months, when increased wall thickness was already evident (Table 1, last line). The extent of cellular hypertrophy did not predict the degree of increased wall thickness seen in *NOS1/3^{-/-}*, consistent with the observation that significant concentric remodelling can occur without corresponding increases in LV mass²⁰.

The current findings demonstrate a new biological paradigm for intracrine NO signalling^{1,6,7,22}, whereby spatial confinement of different NO synthase isoforms allows NO signals to have independent, and even opposite, effects on cardiac phenotype. As such, local regulation of effector molecules is a central mechanism by which NO exerts biological activity^{23–25}. NO can therefore provide exquisite fine-tuning of organ function by recruiting different downstream NO signalling pathways (for example, S-nitrosylation^{26,27} or cGMP production²⁸) within distinct microdomains of the same cell. In addition, we have shown that NO synthase isoforms contribute independently to the maintenance of cardiac architecture, a finding that has important implications for the pathophysiology of congestive heart failure, dilated cardiomyopathy, hypertensive heart disease, and ageing-related cardiac dysfunction. □

Methods

Please see Supplementary Information for details.

Integrated haemodynamic analysis

We studied C57BL/6 transgenic and wild-type mice (2–5 months, Jackson Laboratories). *NOS1^{-/-}* and *NOS3^{-/-}* mice were crossbred to generate a *NOS1/3^{-/-}* phenotype by M. Fishman²⁹. Anaesthetized mice were instrumented with a micromanometer-conductance catheter (SPR-719, Millar Instruments) Isoproterenol was administered via the right jugular vein.

Electrophysiology

[Ca²⁺]_i and sarcomere shortening were recording in fura-2-loaded myocytes using an IonOptix system with an intensified charged coupled device camera and a dual-excitation spectrofluorometer. Myocytes were whole-cell patch-clamped at 37 °C to determine *I*_{CaL}.

Echocardiography

Echocardiography was performed on lightly anaesthetized mice (Sonos 5500 Echocardiogram; Agilent) with a 15-MHz linear transducer. Mean LV wall thickness, LV mass normalized to body weight¹⁹, and relative wall thickness²⁰ were used to index LV hypertrophy, which was concentric among all strains.

Statistical analysis

Data are reported as mean ± s.e.m. Concentration-effect relationships were analysed by two-way repeated-measures analysis of variance (ANOVA) with interaction terms between isoproterenol dose and genotype. Echocardiographic data were analysed with multiple linear regression, including terms for age, genotype, and age–genotype interaction.

Received 24 August; accepted 17 January 2001.

- Fang, M. *et al.* Dexas1: a G protein specifically coupled to neuronal nitric oxide synthase via CAPON. *Neuron* **28**, 183–193 (2000).
- Mery, P. F., Pavoiné, C., Belhassen, L., Pecker, F. & Fischmeister, R. Nitric oxide regulates cardiac Ca²⁺ current. Involvement of cGMP-inhibited and cGMP-stimulated phosphodiesterases through guanylyl cyclase activity. *J. Biol. Chem.* **268**, 26286–26295 (1993).
- Xu, L., Eu, J. P., Meissner, G. & Stamler, J. S. Activation of the cardiac calcium release channel (ryanodine receptor) by poly-S-nitrosylation. *Science* **279**, 234–237 (1998).
- Eu, J. P., Sun, J., Xu, L., Stamler, J. S. & Meissner, G. The skeletal muscle calcium release channel: Coupled O₂ sensor and NO signaling functions. *Cell* **102**, 499–509 (2000).
- Petroff, M. G. *et al.* Endogenous nitric oxide mechanisms mediate the stretch dependence of Ca²⁺ release in cardiomyocytes. *Nature Cell Biol.* **3**, 867–873 (2001).
- Feron, O., Saldana, F., Michel, J. B. & Michel, T. The endothelial nitric-oxide synthase-caveolin regulatory cycle. *J. Biol. Chem.* **273**, 3125–3128 (1998).
- Carcia-Cardena, G. *et al.* Dissecting the interaction between nitric oxide synthase (NOS) and caveolin: Functional significance of the NOS caveolin binding domain *in vivo*. *J. Biol. Chem.* **272**, 25437–25440 (1997).
- Hare, J. M. *et al.* Contribution of caveolin protein abundance to augmented nitric oxide signaling in conscious dogs with pacing-induced heart failure. *Circ. Res.* **86**, 1085–1092 (2000).
- Schwencke, C. *et al.* Compartmentation of cyclic adenosine 3',5'-monophosphate signaling in caveolae. *Mol. Endocrinol.* **13**, 1061–1070 (1999).
- Hare, J. M., Givertz, M. M., Creager, M. A. & Colucci, W. S. Increased sensitivity to nitric oxide synthase inhibition in patients with heart failure: Potentiation of β-adrenergic inotropic responsiveness. *Circulation* **97**, 161–166 (1998).
- Xu, K. Y., Huso, D. L., Dawson, T., Bredt, D. S. & Becker, L. C. NO synthase in cardiac sarcoplasmic reticulum. *Proc. Natl Acad. Sci. USA* **96**, 657–662 (1999).
- Varghese, P. *et al.* β(3)-adrenoceptor deficiency blocks nitric oxide-dependent inhibition of

- myocardial contractility. *J. Clin. Invest.* **106**, 697–703 (2000).
- Hare, J. M., Kass, D. A. & Stamler, J. S. The physiological response to cardiovascular 'orphan' G protein-coupled receptor agonists. *Nature Med.* **5**, 1241–1242 (1999).
- Gyurko, R., Kuhlencordt, P., Fishman, M. C. & Huang, P. L. Modulation of mouse cardiac function *in vivo* by eNOS and ANP. *Am. J. Physiol. Heart Circ. Physiol.* **278**, H971–H981 (2000).
- Cheng, H. J. *et al.* Upregulation of functional β(3)-adrenergic receptor in the failing canine myocardium. *Circ. Res.* **89**, 599–606 (2001).
- Schmidt, A. G. *et al.* Cardiac-specific overexpression of calsequestrin results in left ventricular hypertrophy, depressed force-frequency relation and pulsus alternans *in vivo*. *J. Mol. Cell. Cardiol.* **32**, 1735–1744 (2000).
- Zhai, J. *et al.* Cardiac-specific overexpression of a superinhibitory pentameric phospholamban mutant enhances inhibition of cardiac function *in vivo*. *J. Biol. Chem.* **275**, 10538–10544 (2000).
- Molkentin, J. D. *et al.* A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* **93**, 215–228 (1998).
- Yang, X. P. *et al.* Endothelial nitric oxide gene knockout mice: cardiac phenotypes and the effect of angiotensin-converting enzyme inhibitor on myocardial ischemia/reperfusion injury. *Hypertension* **34**, 24–30 (1999).
- Verdecchia, P. *et al.* Adverse prognostic significance of concentric remodeling of the left ventricle in hypertensive patients with normal left ventricular mass. *J. Am. Coll. Cardiol.* **1925**, 871–878 (1995).
- Topol, E. J., Traill, T. A. & Fortuin, N. J. Hypertensive hypertrophic cardiomyopathy of the elderly. *N. Engl. J. Med.* **312**, 277–283 (1985).
- Kanai, A. J. *et al.* Identification of a neuronal nitric oxide synthase in isolated cardiac mitochondria using electrochemical detection. *Proc. Natl Acad. Sci. USA* **98**, 14126–14131 (2001).
- Kobzik, L., Reid, M. B., Bredt, D. S. & Stamler, J. S. Nitric oxide in skeletal muscle. *Nature* **372**, 546–548 (1994).
- Brennan, J. E. *et al.* Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and α1-syntrophin mediated by PDZ domains. *Cell* **84**, 757–767 (1996).
- Brennan, J. E., Chao, D. S., Xia, H., Aldape, K. & Bredt, D. S. Nitric oxide synthase complexed with dystrophin and absent from skeletal muscle sarcolemma in Duchenne muscular dystrophy. *Cell* **82**, 743–752 (1995).
- Liu, L. *et al.* A metabolic enzyme for S-nitrosothiol conserved from bacteria to humans. *Nature* **410**, 490–494 (2001).
- Stamler, J. S., Lamas, S. & Fang, F. C. Nitrosylation: the prototypic redox-based signaling mechanism. *Cell* **106**, 675–683 (2001).
- Moncada, S. A. The L-arginine-nitric oxide pathway. *N. Engl. J. Med.* **329**, 2002–2012 (1993).
- Son, H. *et al.* Long-term potentiation is reduced in mice that are doubly mutant in endothelial and neuronal nitric oxide synthase. *Cell* **87**, 1015–1023 (1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

This work was supported by grants from the NIH, the American Heart Association, and the American Federation for Aging Research.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M.H. (e-mail: jhare@mail.jhmi.edu).

The AID enzyme induces class switch recombination in fibroblasts

Il-mi Okazaki*, Kazuo Kinoshita*, Masamichi Muramatsu*, Kiyotsugu Yoshikawa*† & Tasuku Honjo*

* Department of Medical Chemistry and Molecular Biology, Graduate School of Medicine; and † Radioisotope Research Center; Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

The switch of the immunoglobulin isotype from IgM to IgG, IgE or IgA is mediated by class switch recombination (CSR). CSR changes the immunoglobulin heavy chain constant region (C_H) gene from C_μ to one of the other C_H genes^{1,2}. Somatic hypermutation introduces massive numbers of point mutations in the immunoglobulin variable (V) region gene, giving rise to immunoglobulin with higher affinity³. Activation-induced cytidine deaminase (AID), a putative RNA-editing cytidine deaminase,

is expressed strictly in activated B cells and is indispensable in both CSR and somatic hypermutation^{4,5}. But the exact function of AID is unknown. Here we show that ectopic expression of AID induces CSR in an artificial switch construct in fibroblasts at a level comparable to that in stimulated B cells. Sequences around recombination junctions in the artificial substrate have features similar to endogenous CSR junctions. Furthermore, AID-induced CSR in fibroblasts is dependent on transcription of the target S region, as shown in endogenous CSR in B cells. The results show that AID is the only B-cell-specific factor required for initiation of the CSR reaction in the activated locus.

CSR is accompanied by the looping-out deletion of DNA segments between the μ switch (S) region and one of the downstream S regions located immediately 5' to each C_H gene^{1,2}. CSR can be divided into three steps: (1) selection of a target S region by transcription from an intron promoter located 5' to each S region ('germline' transcription)^{6–8}; (2) cleavage of the targeted S and S μ regions by a putative DNA-cleaving enzyme; and (3) repair and ligation of the broken DNA ends by the non-homologous end joining (NHEJ) repair system^{9–11}. Signalling through CD40 and cytokine receptors induces germline transcription of a target S region^{6,7} and expression of AID¹², both of which are essential to CSR. AID is involved in the cleaving step of CSR¹³, although it is unknown whether AID itself cleaves DNA or regulates the putative DNA-cleaving enzyme. Expression of AID is restricted to stimulated B lymphocytes, whereas the NHEJ repair system is constitutively expressed in almost all cell types. We have shown that overexpression of AID alone can induce a low level (1–5%) of switching in the endogenous immunoglobulin locus in a murine B lymphoma cell line, CH12F3-2 (ref. 4). This cell line expresses a small amount of germline C α transcripts without stimulation and can switch efficiently from IgM to IgA when stimulated with CD40L and cytokines *in vitro*¹⁴.

We have shown that the combination of an artificial switch substrate and CH12F3-2 cells helps to dissect the molecular mechanism for CSR^{8,15–17}. To examine whether CSR in the artificial construct is also dependent on AID, we used a new artificial CSR substrate, termed SCI(μ , α), in which the gene for the fusion protein hygromycin phosphotransferase–thymidine kinase (HyTK)¹⁸ is inserted between the two transcription units of the S regions (Fig. 1a). SCI(μ , α) contains the S μ and S α regions directed by the elongation factor 1 α promoter (pEF1 α) and a tetracycline-responsive promoter (pTET), respectively. Both S sequences are removed by splicing from each transcript. The coding sequences for the extracellular domain of CD8 α and the sequence for the transmembrane domain of CD8 α (TM) fused with green fluorescent protein (GFP) are separated into two transcription units. The extracellular domain of CD8 α could be anchored on the cell surface only after its fusion with the transmembrane domain by recombination between the two S regions, allowing detection of switched cells by fluorescence-activated cell sorting (FACS).

A CH12F3-2 transfectant with the tetracycline transactivator (tTA)⁸ was stably transfected with SCI(μ , α). Eight clones with a single copy of SCI(μ , α) were selected and stimulated with CD40L, interleukin-4 (IL-4) and transforming growth factor- β (TGF β) in the absence of tetracycline. Surface expression of CD8 α –GFP was detected in about 2–6% of all clones 7 days after stimulation (Fig. 1b and data not shown). This level of switching is comparable to that of the previous construct^{8,15,16}. Surface expression of CD8 α –GFP was dependent on transcription of S regions and stimulation with cytokines as reported previously⁸. CSR efficiency in the endogenous immunoglobulin locus was not augmented by removal of tetracycline from the culture medium, excluding a negative effect of tetracycline on CSR¹⁹ at the concentration employed (50 ng ml^{–1}) (Fig. 1b).

To test whether stimulation can be replaced with exogenous AID expression, we stably transfected one of CH12F3-2-SCI(μ , α) clones

with a plasmid containing AID complementary DNA under the control of pTET. Three independent transfectants showed the appearance of surface CD8 α –GFP⁺ cells without cytokine stimulation but only by removal of tetracycline, which induces both

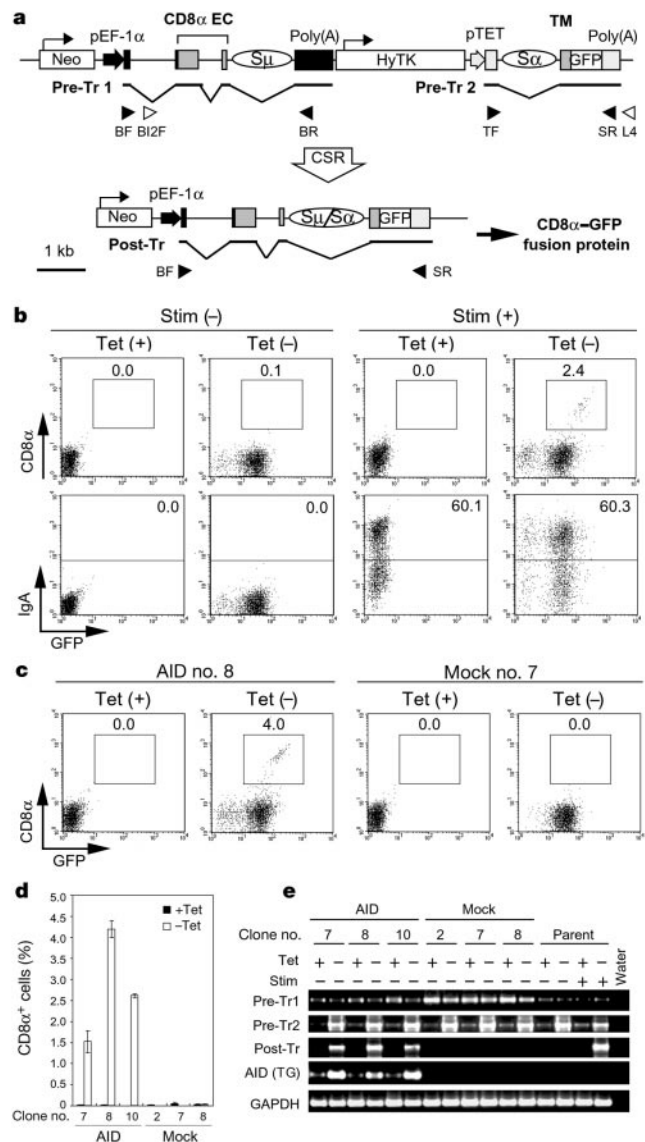


Figure 1 Induction of CSR in an artificial substrate in CH12F3-2 cells by overexpression of AID without stimulation. **a**, Structure of the CSR substrate SCI(μ , α). pEF-1 α , elongation factor 1 α promoter; pTET, tetracycline-responsive promoter; Neo, neomycin resistance gene with the TK promoter; HyTK, gene for the fusion protein hygromycin phosphotransferase–thymidine kinase, with the CMV promoter; Pre-Tr, pre-switch transcripts; Post-Tr, post-switch transcripts; rectangles, exons; arrows, promoters. EC and TM indicate extracellular and transmembrane domains of CD8 α , respectively. Open and filled triangles indicate PCR primers used for genomic DNA and transcripts, respectively. Primer sequences are described in the Methods. **b**, A CH12F3-2 transfectant carrying the tTA and a single and intact copy of the switch substrate SCI(μ , α) was stimulated in the absence or presence of tetracycline for 7 days and analysed for expression of IgA, CD8 α and GFP by FACS. Percentages of cells expressing CD8 α –GFP (top) and IgA (bottom) are indicated. **c**, CH12F3-2-SCI(μ , α) clones transfected with AID expression vector (AID) or the vector alone (mock) under the control of pTET were cultured for 7 days in the absence or presence of tetracycline and analysed for expression of CD8 α and GFP by FACS. **d**, Percentages of CD8 α –GFP⁺ cells in three independent clones of the AID or mock transfectants. **e**, RT-PCR of pre-switch (Pre-Tr1 and 2), post-switch (Post-Tr) and transgenic (TG) AID transcripts in AID-transfected clones. GAPDH is an internal control. Water, without cDNAs. Stim, stimulation with anti-CD40 monoclonal antibody (HM40-3), IL-4 and TGF β .

transcription of $\text{S}\alpha$ and expression of AID (Fig. 1c, d). On day 7, the frequencies of the $\text{CD8}\alpha\text{-GFP}^+$ population in AID-expressing CH12F3-2-SCI(μ,α) cells reached a level comparable to those observed in parental CH12F3-2-SCI(μ,α) cells stimulated for 7 days in the absence of tetracycline (Fig. 1b, d). In control transfectants with the vector alone, the $\text{CD8}\alpha\text{-GFP}^+$ cells did not appear (Fig. 1c). We also confirmed the appearance of transcripts derived from CSR products (Post-Tr) by polymerase chain reaction with reverse transcription (RT-PCR) (Fig. 1e). These results indicate that CSR in the artificial substrate is dependent on AID and that expression of AID alone can confer CSR on an actively transcribed substrate in unstimulated CH12F3-2 cells, suggesting that all the other components required for CSR may be constitutively expressed in B cells.

These results led us to examine whether ectopic expression of AID can induce CSR in other cell types using the artificial construct. We first generated tTA transfectants of a murine fibroblast cell line, NIH3T3, and further transfected one of the transfectants with SCI(μ,α). We then introduced AID by retroviral infection into three independent clones with a single copy of SCI(μ,α). In the absence of tetracycline, surface $\text{CD8}\alpha\text{-GFP}^+$ cells were detected in AID-infected NIH3T3-SCI(μ,α) cells at a level comparable to that

in AID-transfected CH12F3-2-SCI(μ,α) cells (compare Fig. 2a, left, with Fig. 1c). By contrast, a loss-of-function mutant of AID (AIDm-1)²⁰ lacking most of the cytidine deaminase motif did not have such activity. Post-switch transcript was also detected by RT-PCR only in AID-infected cells (Fig. 2b). Furthermore, the $\text{CD8}\alpha\text{-GFP}^+$ population did not appear when cultured in the presence of tetracycline, indicating that these events are dependent on transcription of the $\text{S}\alpha$ region, which is required for CSR *in vivo*^{21–24} (Fig. 2a, right).

Switched cells continued to express $\text{CD8}\alpha\text{-GFP}$ on their surface even after the addition of tetracycline into the culture medium, indicating that expression of GFP was no longer driven by pTET but by pEF-1 α in switched cells (Fig. 2c). Furthermore, when cultured in the presence of hygromycin B, $\text{CD8}\alpha\text{-GFP}^+$ cells died, indicating deletion of the HyTK gene located between the two S regions (Fig. 2c). These results suggest that DNA recombination may have occurred between the S regions in $\text{CD8}\alpha\text{-GFP}^+$ cells, deleting the intervening sequences (Fig. 1a). DNA rearrangements were also confirmed by genomic PCR using BI2F and L4 primers⁸ located 5' to the extracellular domain of $\text{CD8}\alpha$ and 3' to GFP, respectively (Fig. 1a). Genomic DNA from the uninfected and AIDm-1-infected cells gave only the pre-switch band of 10.0 kilobases (kb), whereas

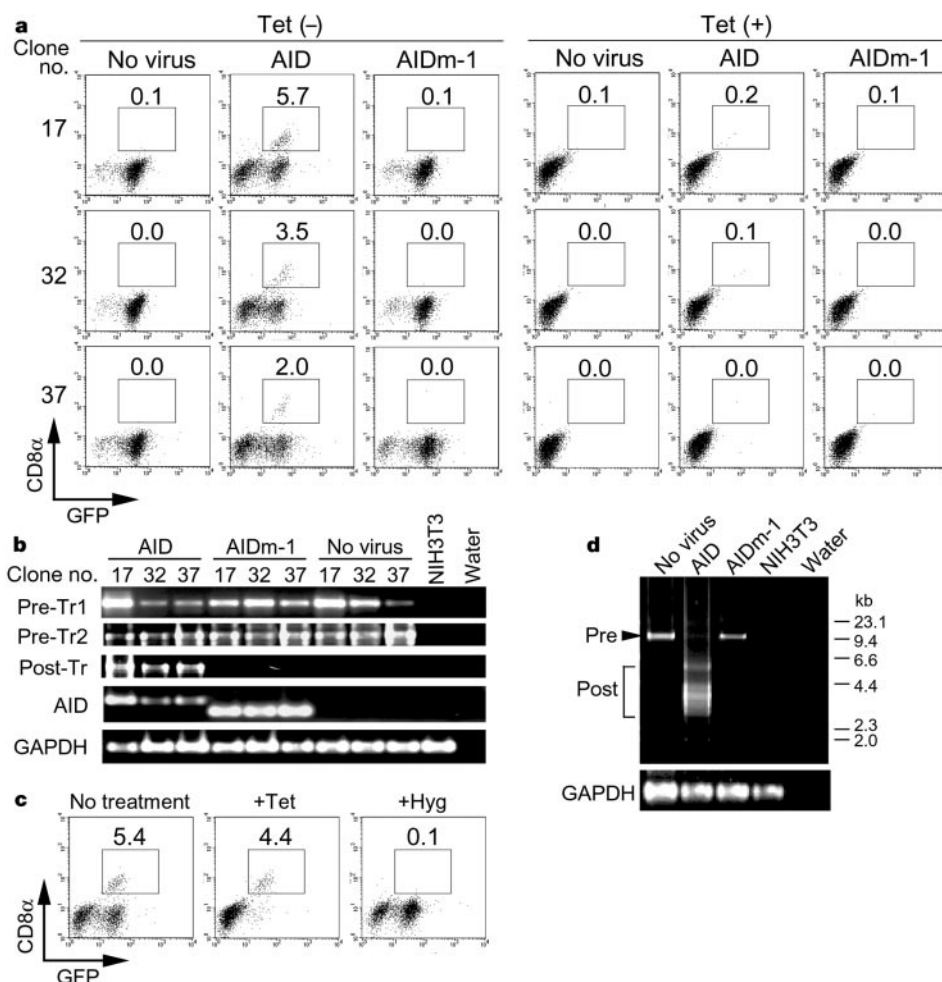


Figure 2 Induction of CSR in the artificial substrate SCI(μ,α) in NIH3T3 cells by expression of AID. **a**, Three independent NIH3T3 transfectant clones carrying tTA and an intact single copy of SCI(μ,α) were infected with AID-expressing retrovirus in the absence or presence of tetracycline. Percentages of $\text{CD8}\alpha\text{-GFP}^+$ cells were analysed by FACS on day 7 and are indicated in each plot. AIDm-1, a loss-of-function mutant of AID²⁰. **b**, RT-PCR of

pre-switch (Pre-Tr1 and 2), post-switch (Post-Tr) and AID transcripts in AID infectants. **c**, AID infectants from clone number 37 were cultured in the presence of tetracycline (middle) or hygromycin B (right). **d**, Genomic PCR of pre- and post-switch products from SCI(μ,α) in AID infectants of NIH3T3-SCI(μ,α) clone number 37.

genomic DNA from AID-infected cells gave, in addition to the faint pre-switch band, heterogeneous smeared products of 3.0–5.5 kb, as expected from heterogeneous CSR between the two S regions (Fig. 2d). Similarly, CSR was induced in another CSR substrate, SCG(μ ,2 α)¹⁶, in EL-4 cells (a murine T-cell line) and L cells (a murine fibroblast cell line) by expression of AID (data not shown). AID overexpression did not induce either germline transcription or CSR in the endogenous immunoglobulin loci in NIH3T3 cells, indicating that AID cannot affect the endogenous immunoglobulin loci that are not accessible (data not shown).

To further confirm that CSR induced in NIH3T3 cells by expression of AID mimics physiological CSR, CD8 α -GFP⁺ clones were obtained by limiting dilution of AID-infected cells. All wells inoculated with a CD8 α -GFP⁺ single cell gave rise to a mixture of CD8 α -GFP⁺ and CD8 α ⁻ GFP⁻ cells (Fig. 3a). The CD8 α ⁻ GFP⁻

population is likely to be derived from the CD8 α -GFP⁺ population during culture, because in each mixture of single-cell progenies, pre-switch transcripts from the two promoters completely disappeared, and post-switch transcripts appeared (Fig. 3b). Also, genomic PCR of each mixture of DNA yielded a discrete post-switch band of 3.0–5.5 kb, and none of these DNAs gave the 10.0-kb pre-switch band, indicating that the same size of DNA segment between two S regions was deleted in each mixture (Fig. 3c). By sequencing these post-switch bands, we determined the recombination junctions. In all clones, recombination occurred between two S regions as expected (Fig. 3d). Furthermore, the features of junction points were similar to those observed in physiological CSR: no consensus sequences around break points, no long homology between two S sequences at the junctions, and frequent mutations in the proximity of the junctions^{2,25}. Recombination junctions distributed in wide ranges of the S μ and S α regions, although break points in the S μ region were biased to the 5' side (Fig. 3e) as reported in the endogenous immunoglobulin loci²⁶. Clone 6 had additional duplications and deletions within the S α region (Fig. 3e), suggesting that this clone might have undergone secondary recombination during culture, again reported previously in the endogenous immunoglobulin locus of B cells^{25,27,28}. These results indicate that AID can induce CSR in the accessible S region in non-lymphoid cells in a manner similar to the endogenous immunoglobulin locus in B cells. We conclude that expression of AID is necessary and sufficient to confer CSR between transcribed S regions in non-lymphoid cells. Therefore, expression of all the components required for the CSR reaction, except AID, are not restricted to activated mature B lymphocytes, but are expressed constitutively and probably ubiquitously.

The CSR efficiency increased dose dependently on the amount of AID viruses used for infection (Fig. 4a). CD8 α -GFP⁺ cells were detectable as early as day 3 and reached the maximum level at day 7

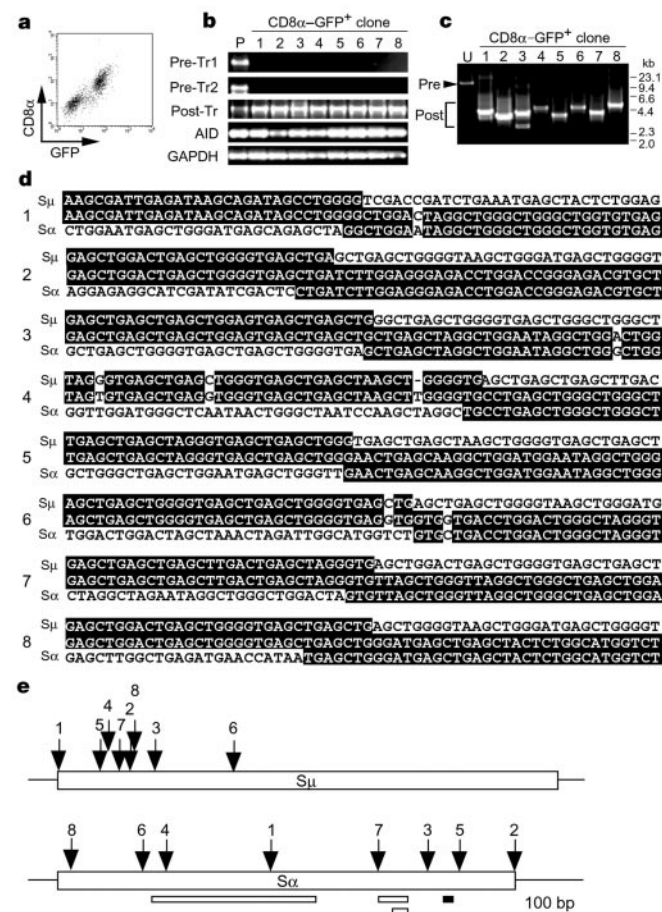


Figure 3 AID-induced CSR in NIH3T3 cells mimics physiological CSR. **a**, FACS profile of a representative post-switch clone 1. **b**, RT-PCR of pre-switch (Pre-Tr1 and 2), post-switch (Post-Tr) and AID transcripts in post-switch clones from three independent AID infectants. **c**, Genomic PCR of pre- and post-switch products from SC1(μ , α) in post-switch clones. **d**, Recombination junctions between S μ and S α regions in post-switch clones. The recombinant sequences (middle) compare with S μ (top) and S α (bottom) sequences. Identical sequences are boxed to represent each break point. Many of the junctions are mapped in 1–4-base homology regions, which inhibits pinpointing of the junction nucleotides. Insertion is shown by a dash. **e**, Distribution of recombination junctions in S μ and S α regions. Arrows show the location of junctions of indicated clones. Open and filled rectangles represent duplications and deletions of clone 6, respectively. Parental transfectants are number 17 for clones 1–3, number 32 for clones 4 and 5, and number 37 for clones 6–8.

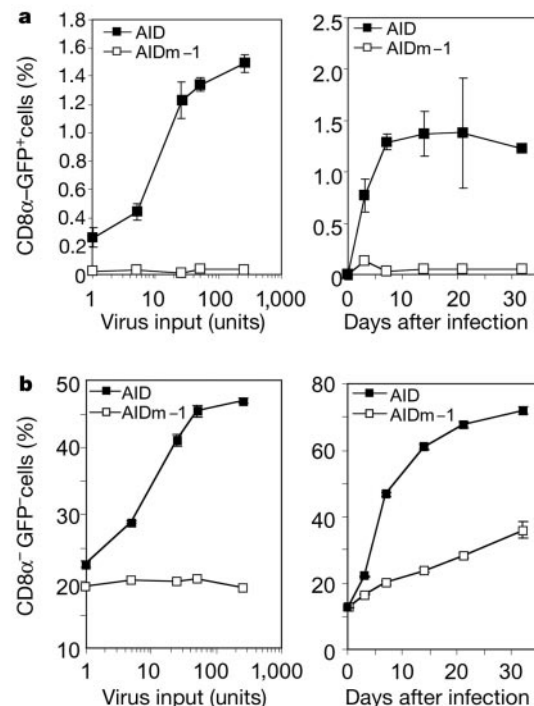


Figure 4 Time course and dose dependency of AID-induced phenotypes in NIH3T3-SC1(μ , α) clone number 37. AID-induced accumulation of CD8 α -GFP⁺ (**a**) and CD8 α -GFP⁻ (**b**) at day 7 with different virus amounts (left) or at a fixed virus titre (50 units) at different days indicated (right). Two other independent clones showed similar profiles.

after infection; the cell numbers then gradually declined (Fig. 4a). Unexpectedly, not only CD8 α -GFP⁺ but also CD8 α ⁻ GFP⁻ cells increased in an AID-dependent manner during culture of NIH3T3-SCI(μ , α) cells (Fig. 4b). The CD8 α ⁻ GFP⁻ cells increased also in CH12F3-2-SCI(μ , α) cells, either by stimulation (Fig. 1b) or by AID expression (Fig. 1c). Because most of the CD8 α ⁻ GFP⁻ cells were resistant to hygromycin B (Fig. 2c) and yet they could also arise from the CD8 α -GFP⁺ cells (Fig. 3a), we assumed that loss of GFP expression might be ascribed to mutations rather than to gross recombination in the GFP gene of the unswitched construct and the switched construct. We amplified TM-GFP fragment sequences in non-rearranged SCI(μ , α) integrated in NIH3T3 cells by nested PCR. Of 22 clones of the fragments amplified from AID-infected cells, 4 had either microduplication or microdeletion in addition to point mutations, and 20 had mutations that appeared to affect GFP fluorescence, judging from a previous structural study²⁹. TM-GFP sequences in AID-infected cells had frequent point mutations (77 of 25,061 base pairs (bp) = 3.1×10^{-3} per bp), whereas those in AIDm-1-infected and uninfected cells had point mutations at the same level with PCR error frequency (4 of 11,671 bp = 3.4×10^{-4} per bp, and 8 of 14,364 bp = 5.6×10^{-4} per bp, respectively). These results suggest that expression of AID alone might also be enough to introduce mutations in non-B cells. Perhaps somatic hypermutation is also induced in fibroblasts by AID expression. Because CD8 α ⁻ GFP⁻ cells slightly increased in AIDm-1-infected cells (Fig. 4b) and in uninfected cells (data not shown) during prolonged culture, silencing may be involved in losing GFP expression, although its contribution seems to be smaller than is mutation by AID expression.

Our findings clearly indicate that AID expression is sufficient to confer CSR in the transcribed—thus accessible—switch substrate in unstimulated B cells (CH12F3-2 cells), T cells (EL-4 cells) and even in non-lymphoid cells (NIH3T3 and L cells). The molecular features of these CSR products are almost the same as those of the physiological CSR products in stimulated B cells. This implies constitutive expression of all the factors required for the CSR reaction, except AID, in non-lymphoid cells and B cells. If so, numerous events triggered by signalling through CD40 and cytokine receptors in B cells are required to induce transcription of target S regions and expression of AID for initiating CSR.

That CSR in the artificial substrate is completely dependent on AID in B cells and fibroblasts strongly supports that our previous findings using artificial substrates reflect the properties of physiological CSR: the absence of primary sequence specificity in the S region¹⁵; requirement of palindromic sequences in S regions¹⁶; and quantitative correlation between efficiencies of S-region transcription and CSR⁸. Although the present finding alone does not allow us to distinguish whether AID is an RNA-editing enzyme², our system will be useful to test this possibility. Furthermore, this system is useful for testing other artificial constructs for AID-dependent CSR. Assuming that a new DNA-cleaving enzyme is generated by RNA-editing activity of AID², one may be able to assay and screen such a product of the AID function using the NIH3T3 transfectants containing the artificial switch construct. Alternatively, AID itself may attack DNA. □

Methods

CSR assay by an artificial substrate

The switch substrate SCI(μ , α) was constructed by inserting FS μ -4 and FS α -2 fragments, core S μ and S α regions of mouse, respectively, into SCI(0,0), which was derived from SCT(0,0) by deleting the SmaI site in pTET and inserting the HyTK gene¹⁸. The linearized plasmid constructs were introduced into CH12F3-2 cells and NIH3T3 cells by electroporation and Lipofectamine (Gibco BRL), respectively. The integration status of SCI(μ , α) in G418-resistant clones was examined by Southern blotting, and clones with a single and intact copy were used for further analyses.

CH12F3-2 cells were cultured as described¹⁴. CSR stimulation of CH12F3-2 was performed as described¹⁷. NIH3T3 cells were cultured in DMEM supplemented with 10% heat-inactivated fetal calf serum, 2 mM L-glutamine and antibiotics. CSR in

transfectants was assessed by expression of CD8 α -GFP on the surface. Cells were stained with allophycocyanin (APC)-conjugated anti-mouse CD8 α monoclonal antibodies (Pharmingen) and analysed on a FACScalibur with CellQuest software (Becton Dickinson). Dead cells were excluded from the analysis by forward- and side-scatter intensity and propidium iodide gating.

RT-PCR and genomic PCR

Total RNA was isolated using TRIzol reagent (Gibco BRL), according to the manufacturer's instructions, and reverse transcribed using oligo(dT) primers and SuperScript II reverse transcriptase (Gibco BRL). Pre- and post-switch transcripts were amplified as described previously⁸ by 30 cycles of PCR using the primers (BF, BR, TF and SR) described, except that Pre-Tr1 was amplified for 35 cycles. AID transcripts from pTETbsr-AID transgene and pMX-mAID transgene loci were amplified by 30 cycles of PCR using the following primer pairs: TF (specific for the pTET) and AID-R2 (5'-GCCCTGGCGGTCTTCACAGAAGTAG-3'); and AID-S1 (5'-CGGCATGAGACCTACCTCTGCTAC-3') and AID-R2, respectively. As an internal control, GAPDH transcripts were amplified by 20 cycles of PCR as described⁸.

For genomic PCR, genomic DNA was isolated by a standard procedure. PCR was performed using LA Taq polymerase (TaKaRa) as described previously⁸. The PCR products were sequenced either directly using the primers described¹⁶, or after digestion and subcloning into pBluescript using T3 and T7 primers.

Exogenous expression of AID

A tetracycline-regulated AID expression vector was constructed by cloning AID cDNA into pTETbsr, which was generated by cloning the blasticidin S-resistant gene cassette (bsr) of pSV2bsr (Kaken Seiyaku) into pTet-Splice (Gibco BRL). AIDm-1 was described previously²⁰. For retroviral infection, AID and AIDm-1 cDNAs were subcloned into pMX vector³⁰. Recombinant retroviruses were produced as described previously²⁰. Virus titre was determined by activity to rescue *in vitro* switching in AID^{-/-} primary B cells. One unit of AID virus was defined to give 1% of switching in AID^{-/-} primary B cells. The infection efficiency of the pMX-AID retrovirus supernatant to NIH3T3 cells was estimated to be around 30% at 1 unit, 70% at 5 units and >95% at concentrations higher than 25 units, as judged from the infection efficiency of the pMX-AID-IRES-GFP retrovirus, which had an equivalent titre in AID^{-/-} primary B cells. The amount of the pMX-AIDm-1 retrovirus supernatant was estimated by comparison of the amount of transcripts with that of the pMX-AID retrovirus supernatant in NIH3T3. Retroviral infection was performed as follows. NIH3T3 cells were seeded at a density of 5×10^4 cells per six-well dish and incubated for 16 h. The medium was then replaced with fresh medium containing retroviruses supplemented with $8 \mu\text{g ml}^{-1}$ polybrene (Sigma), centrifuged at 1,830 g for 1 h at 32°C, and incubated at 37°C. AID and AIDm-1 expression was confirmed by western blotting using Flag-tagged constructs.

Received 14 December 2001; accepted 8 January 2002.

Published online 3 March 2002, DOI 10.1038/nature727.

1. Honjo, T., Kinoshita, K. & Muramatsu, M. Molecular mechanism of class switch recombination: linkage with somatic hypermutation. *Annu. Rev. Immunol.* **20**, 165–196 (2002); published online 4 October 2001 (DOI 10.1146/annurev.immunol.20.081701.141316v4).
2. Kinoshita, K. & Honjo, T. Linking class-switch recombination with somatic hypermutation. *Nature Rev. Mol. Cell Biol.* **2**, 493–503 (2001).
3. Neuberger, M. S. & Milstein, C. Somatic hypermutation. *Curr. Opin. Immunol.* **7**, 248–254 (1995).
4. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
5. Revy, P. *et al.* Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**, 565–575 (2000).
6. Stavnezer, N. J. & Sirlin, S. Specificity of immunoglobulin heavy chain switch correlates with activity of germ-line heavy chain genes prior to switching. *EMBO J.* **5**, 95–102 (1986).
7. Yancopoulos, G. D. *et al.* Secondary genomic rearrangement events in pre-B cells: VHDJH replacement by a LINE-1 sequence and directed class switching. *EMBO J.* **5**, 3259–3266 (1986).
8. Lee, C. G. *et al.* Quantitative regulation of class switch recombination by switch region transcription. *J. Exp. Med.* **194**, 365–374 (2001).
9. Rolink, A., Melchers, F. & Andersson, J. The SCID but not the RAG-2 gene product is required for S μ -Se heavy chain class switching. *Immunity* **5**, 319–330 (1996).
10. Casellas, R. *et al.* Ku80 is required for immunoglobulin isotype switching. *EMBO J.* **17**, 2404–2411 (1998).
11. Manis, J. P. *et al.* Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. *J. Exp. Med.* **187**, 2081–2089 (1998).
12. Muramatsu, M. *et al.* Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* **274**, 18470–18476 (1999).
13. Petersen, S. *et al.* AID is required to initiate Nbs1/γ-H2AX focus formation and mutations at sites of class switching. *Nature* **414**, 660–665 (2001).
14. Nakamura, M. *et al.* High frequency class switching of an IgM⁺ B lymphoma clone CH12F3 to IgA⁺ cells. *Int. Immunol.* **8**, 193–201 (1996).
15. Kinoshita, K., Tashiro, J., Tomita, S., Lee, C. G. & Honjo, T. Target specificity of immunoglobulin class switch recombination is not determined by nucleotide sequences of S regions. *Immunity* **9**, 849–858 (1998).
16. Tashiro, J., Kinoshita, K. & Honjo, T. Palindromic but not G-rich sequences are targets of class switch recombination. *Int. Immunol.* **13**, 495–505 (2001).
17. Chen, X., Kinoshita, K. & Honjo, T. Variable deletion and duplication at recombination junction ends: Implication for staggered double-strand cleavage in class-switch recombination. *Proc. Natl Acad. Sci. USA* **98**, 13860–13865 (2001).
18. Lupton, S. D., Brunton, L. L., Kalberg, V. A. & Overell, R. W. Dominant positive and negative selection using a hygromycin phosphotransferase–thymidine kinase fusion gene. *Mol. Cell Biol.* **11**,

3374–3378 (1991).

19. Kuzin, I. I. *et al.* Tetracyclines inhibit activated B cell function. *Int. Immunol.* **13**, 921–931 (2001).
20. Fagarasan, S., Kinoshita, K., Muramatsu, M., Ikuta, K. & Honjo, T. *In situ* class switching and differentiation to IgA-producing cells in the gut lamina propria. *Nature* **413**, 639–643 (2001).
21. Gu, H., Zou, Y. R. & Rajewsky, K. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* **73**, 1155–1164 (1993).
22. Jung, S., Rajewsky, K. & Radbruch, A. Shutdown of class switch recombination by deletion of a switch region control element. *Science* **259**, 984–987 (1993).
23. Xu, L. *et al.* Replacement of germ-line e promoter by gene targeting alters control of immunoglobulin heavy chain class switching. *Proc. Natl Acad. Sci. USA* **90**, 3705–3709 (1993).
24. Zhang, J., Bottaro, A., Li, S., Stewart, V. & Alt, F. W. A selective defect in IgG2b switching as a result of targeted mutation of the I γ 2b promoter and exon. *EMBO J.* **12**, 3529–3537 (1993).
25. Dunnick, W., Hertz, G. Z., Scappino, L. & Grtzmacher, C. DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res.* **21**, 365–372 (1993).
26. Iwasato, T., Arakawa, H., Shimizu, A., Honjo, T. & Yamagishi, H. Biased distribution of recombination sites within S regions upon immunoglobulin class switch recombination induced by transforming growth factor β and lipopolysaccharide. *J. Exp. Med.* **175**, 1539–1546 (1992).
27. Obata, M. *et al.* Structure of a rearranged γ 1 chain gene and its implication to immunoglobulin class-switch mechanism. *Proc. Natl Acad. Sci. USA* **78**, 2437–2441 (1981).
28. Winter, E., Krawinkel, U. & Radbruch, A. Directed Ig class switch recombination in activated murine B cells. *EMBO J.* **6**, 1663–1671 (1987).
29. Cubitt, A. B. *et al.* Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **20**, 448–455 (1995).
30. Kitamura, T. New experimental approaches in retrovirus-mediated expression screening. *Int. J. Hematol.* **67**, 351–359 (1998).

Acknowledgements

We are grateful to H. Yagita and Targeted Genetics Co. for providing the HM40-3 monoclonal antibody and the tgCMV/HyTK plasmid, respectively. We thank Y. Sakakibara, K. Ikuta, H. Nagaoka, S. Fagarasan and T. Okazaki for their critical reading of the manuscript; T. Muto, T. Eto and T. Doi for discussions; Y. Tabuchi, T. Toyoshima and K. Yurimoto for their technical assistance; and T. Nishikawa for secretarial help. This work was supported by a Center of Excellence Grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan. I.O. is a research fellow of the Japan Society for the Promotion of Science.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.H.
(e-mail: honjo@mfour.med.kyoto-u.ac.jp).

TNF-RII and c-IAP1 mediate ubiquitination and degradation of TRAF2

Xiaoming Li, Yili Yang & Jonathan D. Ashwell

Laboratory of Immune Cell Biology, National Cancer Institute,
National Institutes of Health, Bethesda, Maryland 20892, USA

Tumour necrosis factor- α (TNF- α) is a proinflammatory mediator that exerts its biological functions by binding two TNF receptors (TNF-RI and TNF-RII), which initiate biological responses by interacting with adaptor and signalling proteins. Among the signalling components that associate with TNF receptors are members of the TNF-R-associated factor (TRAF) family^{1,2}. TRAF2 is required for TNF- α -mediated activation of c-Jun N-terminal kinase (JNK), contributes to activation of NF- κ B, and mediates anti-apoptotic signals^{3,4}. TNF-RI and TNF-RII signalling complexes also contain the anti-apoptotic ('inhibitor of apoptosis') molecules c-IAP1 and c-IAP2 (refs 5, 6), which also have RING domain-dependent ubiquitin protein ligase (E3) activity⁷. The function of IAPs in TNF-R signalling is unknown. Here we show that binding of TNF- α to TNF-RII induces ubiquitination and proteasomal degradation of TRAF2.

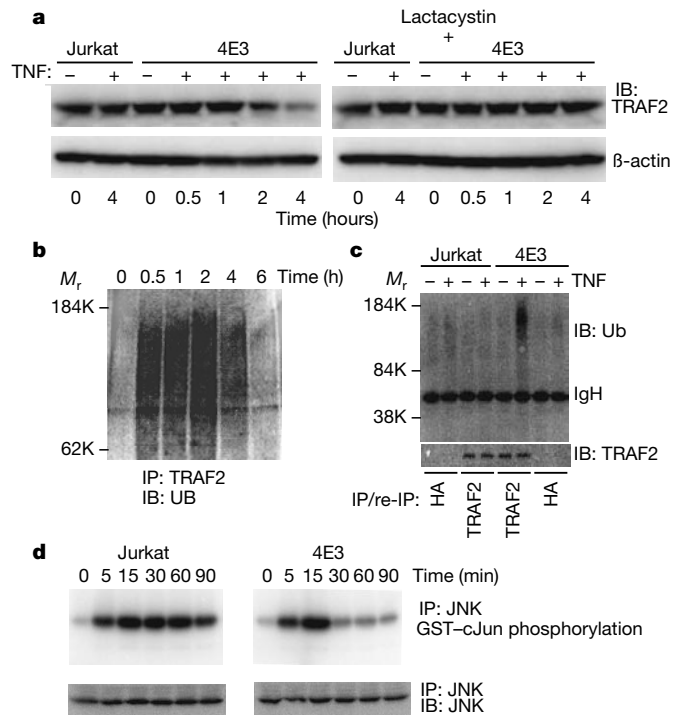


Figure 1 TNF- α induces ubiquitination and proteasome-mediated degradation of TRAF2. **a**, Jurkat and 4E3 cells were cultured in medium (left panel) or pretreated with lactacystin (50 μ M) (right panel) for 60 min and stimulated with TNF- α (20 ng ml⁻¹) for the indicated times. Expression of TRAF2 was detected by immunoblotting (IB). **b**, 4E3 cells were pretreated with MG-132 (40 μ M) for 30 min and stimulated with TNF- α (50 ng ml⁻¹) for the indicated times, at which the cells were lysed, immunoprecipitated (IP) with anti-TRAF2, and immunoblotted with anti-ubiquitin (Ub). **c**, Jurkat and 4E3 cells were pretreated with MG-132 and stimulated with TNF- α as in **b**. After 2 h the cells were lysed and immunoprecipitation was performed with anti-TRAF2 or anti-HA (specificity control). The immunoprecipitated material was eluted from the beads by heating in the presence of 0.5% SDS, and re-immunoprecipitated (re-IP) with the same antibodies. After separation by SDS-PAGE and transfer, the membranes were blotted with anti-ubiquitin (top) or anti-TRAF-2 (bottom). IgH indicates immunoglobulin heavy chain. **d**, Jurkat and 4E3 cells were stimulated with TNF- α for the indicated times. Cell lysates were subjected to immunoprecipitation with anti-JNK antibody. Half of the immunoprecipitated material was used for a kinase assay (upper panel) and half was used for immunoblotting with anti-JNK (lower panels).

Although c-IAP1 bound TRAF2 and TRAF1 *in vitro*, it ubiquitinated only TRAF2. Expression of wild-type c-IAP1, but not an E3-defective mutant, resulted in TRAF2 ubiquitination and degradation. Moreover, E3-defective c-IAP1 prevented TNF- α -induced TRAF2 degradation and inhibited apoptosis. These findings identify a physiologic role for c-IAP1 and define a mechanism by which TNF-RII-regulated ubiquitin protein ligase activity can potentiate TNF-induced apoptosis.

TNF- α induces apoptosis mainly through TNF-RI, which contains an intracellular death domain that recruits death effector adaptor molecules like TRADD and FADD, with the subsequent activation of a caspase cascade⁸. Given that TNF-RII does not contain a death domain and that NF- κ B and the IAPs have anti-apoptotic activity^{9–12}, it is paradoxical that signalling via TNF-RII potentiates the pro-apoptotic effects of concomitant TNF-RI signals^{13,14}. Signalling through TNF-RII results in decreased TRAF2 recovery in Triton X-100 detergent lysates. It was initially suggested that this is due to calpain-like protease activity¹⁵, although in a subsequent study inhibition of calpain activity did not prevent the loss of TRAF2 (ref. 16). More recently it was shown that signalling

through the intracellular domain of some TNF-R family members causes redistribution of TRAF2 to a Triton X-100-insoluble compartment¹⁷. To explore the mechanism of TRAF2 protein down-regulation, Jurkat T cells (which bear TNF-RI but very little TNF-RII) and 4E3 cells (Jurkat T cells stably transfected with TNF-RII) were stimulated with TNF- α , and at different times the cells were lysed in RIPA buffer, which leaves relatively little insoluble material (Fig. 1a). Stimulation of the parent Jurkat cell line had no effect on TRAF2 levels. However, as previously reported after solubilization with Triton X-100 (ref. 16), TNF- α stimulation of 4E3 cells resulted in the progressive loss of TRAF2. No TRAF2 was found in the small amount of RIPA-insoluble material from 4E3 cells, excluding protein partitioning as a cause for the decreased recovery (data not shown). To test the possibility that TRAF2 was being degraded in proteasomes, Jurkat and 4E3 cells were stimulated with TNF- α in the presence of the proteasome inhibitor lactacystin. In this case, no decrease in TRAF2 was observed. The TNF- α -induced decrease in TRAF2 was not prevented by ZVAD (a broad-spectrum caspase inhibitor that prevents apoptosis) or the calpain inhibitor E64d (data not shown). Because proteins are targeted for proteasome-mediated degradation by modification with polyubiquitin chains, 4E3 cells were stimulated with TNF- α in the presence of the proteasome inhibitor MG-132, to allow the accumulation of ubiquitinated species. Within 30 min TNF- α induced a large increase in the amount of polyubiquitinated material immunoprecipitated with anti-TRAF2, with the levels falling 4–6 h after stimulation

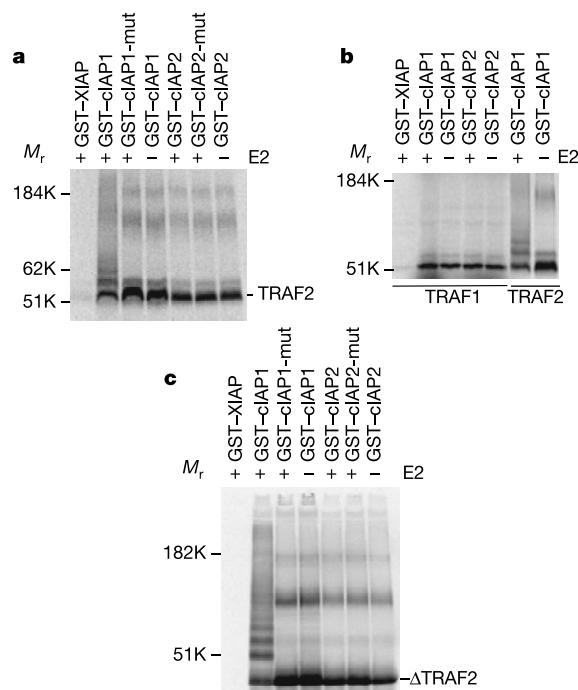


Figure 2 Ubiquitination of TRAFs by IAPs *in vitro*. **a**, c-IAP1 but not c-IAP2 ubiquitinates TRAF2. *In vitro* translated and ³⁵S-labelled TRAF2 was incubated with beads coated with GST-IAP fusion proteins as indicated for 2 h at 4 °C. The beads were washed and subjected to an *in vitro* ubiquitination assay. The bound material was eluted and resolved by SDS-PAGE. One representative experiment of five is shown. **b**, TRAF1 is not a substrate for either c-IAP1 or c-IAP2. ³⁵S-labelled TRAF1 or TRAF2 was incubated with beads coated with the indicated GST-IAP fusion proteins and *in vitro* ubiquitination was performed. Ubiquitination of TRAF2 was used as a positive control. One representative experiment of three is shown. **c**, The RING domain of TRAF2 is dispensable for its ubiquitination by c-IAP1. TRAF2 and ΔTRAF2, which lacks the N-terminal RING domain, were subjected to an *in vitro* ubiquitination assay as above. One representative experiment of three is shown.

(Fig. 1b). To address the possibility that the ubiquitinated material might represent proteins co-precipitated with TRAF2 rather than TRAF2 itself, a reprecipitation experiment was performed (Fig. 1c). Jurkat and 4E3 cells were stimulated with TNF- α and after 2 h the

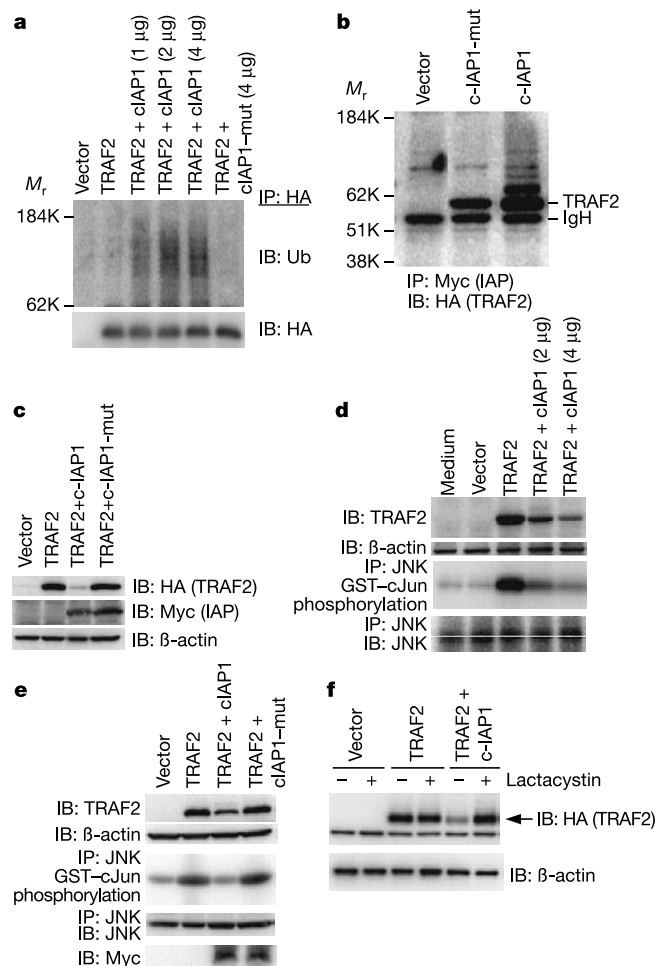
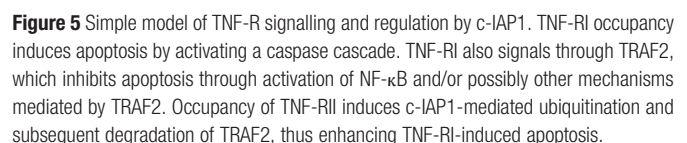
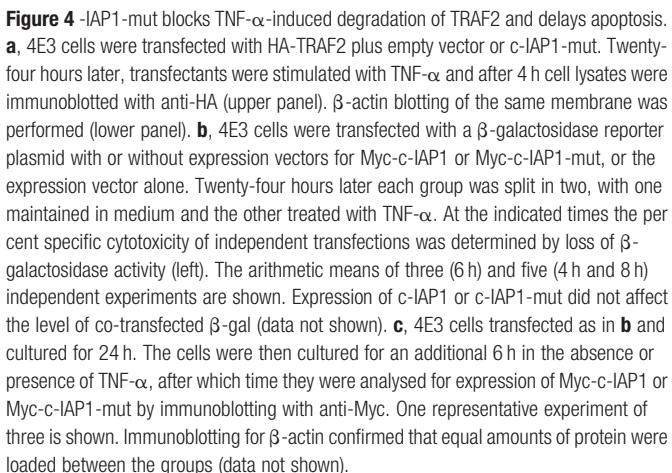


Figure 3 Ubiquitination and proteasome-mediated degradation of TRAF2 by c-IAP1 *in vivo*. **a**, 293 cells were transfected with indicating amounts of expression vectors encoding HA-tagged TRAF2 and Myc-tagged c-IAP1 or c-IAP1-mut (E3-inactive). After 24 h cells were harvested and lysates immunoprecipitated with anti-HA, resolved by SDS-PAGE, and immunoblotted with anti-ubiquitin (upper panel) or anti-HA (lower panel). **b**, 293 cells were transfected with 1 μg of HA-TRAF2 with or without 4 μg of Myc-tagged c-IAP1 or c-IAP-mut. After 20 h, ubiquitin aldehyde (3 μM) was added to inhibit deubiquitination, and 4 h later the cells were lysed, immunoprecipitated with anti-Myc, and immunoblotted with anti-HA to detect c-IAP1-associated HA-TRAF2. **c**, 293 cells were transfected with 1 μg of HA-TRAF2 with or without 4 μg of Myc-tagged c-IAP1 or c-IAP1-mut. After 36 h the samples were split and immunoblotted with anti-HA (to detect TRAF2) or anti-Myc (to detect the IAPs). The membrane used for the anti-HA blot was stripped and reblotted with anti-β-actin antibodies. Results from one of four representative experiments are shown. **d**, 293 cells were transfected with the indicated expression vectors. After lysis, a portion of the material was used for immunoblotting with anti-HA for HA-TRAF2 expression and anti-β-actin, and the rest was subjected to immunoprecipitation with anti-JNK antibodies followed by a JNK kinase assay. The immunoprecipitate was also immunoblotted with anti-JNK. **e**, 293 cells were transfected with the indicated expression vectors and analysed as in **d**, except that anti-Myc was used to detect the transfected IAPs. **f**, 293 cells were transfected with expression vectors encoding HA-TRAF2 and Myc-tagged c-IAP1 as indicated. Twenty-four hours after transfection, lactacystin (50 μM) was added to the medium and 8 h later expression of HA-TRAF2 was assessed by immunoblotting with anti-HA antibodies. β-actin levels were simultaneously measured. One representative experiment of five is shown.

To address the functional importance of c-IAP1-mediated TRAF2 ubiquitination in TNF- α signalling, attempts were made to use E3-defective c-IAP1 as a dominant negative. 4E3 cells were



transfected with HA-TRAF2 with or without E3-defective c-IAP1 and treated with TNF- α . Notably, the TNF- α -induced degradation of TRAF2 was prevented by co-expression of the c-IAP1 RING mutant (Fig. 4a). The E3-inactive c-IAP1 mutant also rescued TRAF2 protein expression in 293 cells and HeLa cells transfected with TNF-RII and treated with TNF- α (data not shown). TRAF2 protects cells from TNF- α -induced death^{22,23}, and thus the prevention of TRAF2 degradation by the E3-defective c-IAP1 might be expected to have a similar effect. Whereas TNF- α did not induce the death of Jurkat cells (ref. 16 and our unpublished observations), TNF- α did kill 4E3 cells (Fig. 4b). Although in some instances c-IAP1 has anti-apoptotic activity by virtue of its BIR-containing amino-terminal region, it did not protect 4E3 cells from TNF- α -induced apoptosis. In fact, overexpression of c-IAP1 had a small but reproducible enhancing effect on TNF cytotoxicity, especially at early time points. The E3-inactive c-IAP1 mutant, on the other hand, substantially inhibited TNF- α -induced apoptosis up to 8 h. The amount of cell death converged after this time, being similar among the groups by 12–16 h (data not shown). The levels of transfected c-IAP1 and c-IAP1-mut (RING mutant) in the 4E3 cells were similar and, unlike TRAF2, treatment with TNF- α did not cause an appreciable change in c-IAP1 or c-IAP1-mut levels (Fig. 4c). Thus, E3-inactive c-IAP1 appears to act in a dominant-negative fashion to prevent TNF-RII-mediated and TNF- α -induced down-regulation of TRAF2 and delay cell death. In preliminary studies we have found that unlike c-IAP1, c-IAP2 inhibits TNF- α -induced apoptosis of 4E3 cells (see Supplementary Information), consistent with its failure to ubiquitinate TRAF2 *in vitro*. That the enhancement of TNF-induced death by c-IAP1 was small presumably indicates that the level of endogenous c-IAP1 in 4E3 cells is not limiting for TRAF2 ubiquitination and degradation (see Fig. 1).

TRAFs are critical mediators of the cell activation and anti-apoptotic functions of the TNF-R superfamily. In particular, TRAF2 is important in protection from apoptosis, as is clear from the observation that cells from TRAF2-deficient mice are overly sensitive to TNF- α -induced apoptosis⁴. Although TNF-RII couples to the pro-survival TRAF2/NF- κ B signalling pathway, cells from mice deficient in TRNF-II are actually less sensitive to TNF- α -induced apoptosis²⁴. Furthermore, TNF-RII is required for TNF- α -induced apoptosis of normal T cells^{13,14}, and the binding of TRAF2 to TNF-RII is required for the potentiation of TNF-RI killing^{22,23}. The data in the present study support a mechanism in which TNF-RII occupancy causes the ubiquitination of TRAF2 by c-IAP1 (Fig. 5). It is also possible that there are other as yet unidentified anti-apoptotic c-IAP1 substrates. We note that RING-less c-IAP1 was effective, while the full-length molecule was not, in preventing apoptosis induced by Sindbis virus, and that a carboxy-terminal (RING-containing) fragment of c-IAP1 actually induced apoptosis when expressed in cell lines²⁵. In addition, other non-death-domain members of the TNF-R superfamily, such as CD30 and CD40, may share the ability to cause TRAF2 degradation, and in the latter case this has been reported to be proteasome-dependent^{26,27}. Finally, mice lacking TNF-RII have enhanced TNF-RI-dependent inflammatory responses²⁸, consistent with a role for TNF-RII in decreasing TRAF2 levels. The results in this report indicate that the ubiquitin protein ligase c-IAP1 dynamically regulates the level and therefore function of TRAF2 in response to TNF-RII occupancy, and paradoxically can play a pro-apoptotic role in TNF- α signalling. □

Methods

Cell lines and reagents

Human embryonic kidney fibroblast 293 (293 cells), HeLa and Jurkat cells were obtained from the American Type Culture Collection. 4E3 cells were generated by stable transfection of Jurkat cells with human TNF-RII cDNA¹⁶, provided by M. Lenardo. Constructs encoding GST-XIAP, GST-c-IAP1, and their mutant forms were generated as described⁷. GST-c-IAP2 and GST-c-IAP2-mut (encoding c-IAP2 residues 1–382, and thus lacking the C-terminal RING domain) were made by cloning the corresponding cDNAs into pGEX4T-2 with *Bam*HI and *Not*I sites. Plasmids used for *in vitro* translation

and/or *in vivo* expression of human TRAFs were pBluescript KS(-)-TRAF1, pBluescript KS(-)-TRAF2, pCDNA3- Δ TRAF2, and pCDNA3-HA-tagged TRAF2 (provided by C. Duckett). Human c-IAP1, c-IAP1(H588A) (c-IAP1-mut), XIAP, and XIAP His(467) (XIAP-mut) were Myc-tagged at the 3' end and cloned into pCDNA3 as described⁷. pCDNA3-TNF-RII cDNA plasmid was provided by P. Scheurich. The pCMV.SPORT- β gal reporter plasmid was purchased from GIBCO/BRL. Wheat E1 and the human E2 UbcH5B were expressed in *Escherichia coli*⁷. Ubiquitin was purchased from Sigma. To obtain radiolabelled ubiquitin, a GST-ubiquitin fusion protein (the construct provided by A. Weissman) was expressed in *E. coli*, and isolated protein phosphorylated with ³²P- γ -ATP. The GST-c-Jun (1–79) construct was provided by Z. Liu. Anti-Myc and anti-HA (mouse, immunoglobulin- γ 1) monoclonal antibodies were derived from mouse hybridoma cell lines 12CA5 and 9E10, respectively. The anti-TRAF2 rabbit antiserum (C-20), and anti-TRAF2 monoclonal antibody (mouse, IgG1), an anti-ubiquitin monoclonal antibody (mouse, IgG1), and anti-JNK1 monoclonal antibodies were purchased from Santa Cruz Biotechnology. MG-132, lactacystin and ubiquitin aldehyde were purchased from CalBiochem. Human TNF- α was purchased from R&D Systems.

Cell cultures and transient transfections

293 cells and HeLa cells were grown in DMEM (BioSource) containing 10% heat-inactivated fetal calf serum, antibiotics, 4 mM glutamine, and 5×10^{-5} M β -mercaptoethanol. Jurkat and 4E3 cells were cultured in RPMI-1640 (BioSource) supplemented as above. Cells were transfected with the indicated expression vectors using Lipofectamine 2000 (GIBCO/BRL) according to the manufacturer's instructions.

Immunoprecipitation and immunoblotting

Cells were lysed in RIPA buffer (1 $\times$ PBS, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS) containing 0.5 mM AEBSF, leupeptin (20 μ g ml⁻¹), and aprotinin (20 μ g ml⁻¹) on ice. In some experiments the lysate was resolved on 10% SDS-PAGE, in others immunoprecipitation with the indicated antibodies was performed and, after boiling the beads, the eluted material was resolved by 8% or 10% SDS-PAGE. After transfer to nitrocellulose (Novex), the membranes were blotted with the indicated antibodies. For most experiments, blots were developed with ¹²⁵I-protein A (ICN) and visualized with a Storm PhosphorImager (Molecular Dynamics). In some experiments, visualization of immunoblots was performed by enhanced chemiluminescence (Amersham Pharmacia) according to the manufacturer's instructions.

In vitro ubiquitination assay

GST-IAP fusion proteins were expressed in *E. coli* BL-21 (DE3) (Novagen), and purified from bacterial lysates with glutathione Sepharose 4B beads (Amersham Pharmacia). Putative IAP substrates were translated *in vitro* and metabolically labelled with ³⁵S-methionine using the TNT kit (Promega). The labelled substrates were incubated with GST-IAP fusion proteins in GST-binding buffer (120 mM NaCl, 10% glycerol, 1% Triton X-100, 50 mM Tris, pH 7.5) at 4 °C for 2 h, after which time the beads were washed and the material bound to the GST fusion proteins was incubated in ubiquitination reaction buffer (50 mM Tris, pH 7.4, 0.2 mM ATP, 0.5 mM MgCl₂, 0.1 mM dithiothreitol (DTT), 1 mM creatine phosphate, 15 units of creatine phosphokinase) containing E1 (20 ng), E2 (20 ng), and 1 μ g of unlabelled ubiquitin at 30 °C²⁹ for 1 h. The eluted material resolved by SDS-PAGE and visualized with a Storm PhosphorImager.

Determination of JNK activity

Activation of endogenous JNK was evaluated using GST-c-Jun (1–79) as a specific substrate. Briefly, JNK was immunoprecipitated from the lysates of Jurkat or 4E3 cells stimulated with TNF- α for the indicated times or 293 cells that had been transfected with HA-TRAF2 and the indicated amounts of Myc-tagged c-IAP1, and kinase reactions carried out at 30 °C for 30 min. The samples were boiled and the eluted material was resolved by 12.5% SDS-PAGE and visualized with a Storm PhosphorImager.

Quantitation of cell death

Death of transfected 4E3 cells was determined by release of β -galactosidase as described³⁰. Briefly, 4E3 cells were transfected with pCMV.SPORT- β gal reporter plasmid (0.25 μ g) with or without c-IAP1 (2.5 μ g) or c-IAP1-mut (2.5 μ g) expression vectors. After 24 h, the cells were split into two groups, one cultured without and one cultured with TNF- α (50 ng ml⁻¹), and at the indicated times the cells were lysed and β -galactosidase activity was measured. TNF- α had no effect on β -galactosidase levels when the cells were treated in the presence of Z-Val-Ala-Asp(O-methyl)-fluoromethylketone (ZVAD; Enzyme Systems Products) to prevent apoptosis (data not shown). Cell death was calculated with the formula: per cent specific cytotoxicity = $100 - [(\text{experimental } \beta\text{-gal/medium } \beta\text{-gal}) \times 100]$.

Received 30 October; accepted 9 January 2001.

1. Rothe, M., Pan, M. G., Henzel, W. J., Ayres, T. M. & Goeddel, D. V. The TNFR2-TRAF signaling complex contains two novel proteins related to baculoviral inhibitor of apoptosis proteins. *Cell* **83**, 1243–1252 (1995).
2. Hsu, H., Shu, H. B., Pan, M. G. & Goeddel, D. V. TRADD-TRAF2 and TRADD-FADD interactions define two distinct TNF receptor 1 signal transduction pathways. *Cell* **84**, 299–308 (1996).
3. Lee, S. Y. *et al.* TRAF2 is essential for JNK but not NF- κ B activation and regulates lymphocyte proliferation and survival. *Immunity* **7**, 703–713 (1997).
4. Yeh, W. C. *et al.* Early lethality, functional NF- κ B activation, and increased sensitivity to TNF-induced cell death in TRAF2-deficient mice. *Immunity* **7**, 715–725 (1997).
5. Shu, H. B., Takeuchi, M. & Goeddel, D. V. The tumor necrosis factor receptor 2 signal transducers TRAF2 and c-IAP1 are components of the tumor necrosis factor receptor 1 signaling complex. *Proc. Natl Acad. Sci. USA* **93**, 13973–13978 (1996).

6. Chan, F. K., Siegel, M. R. & Lenardo, J. M. Signaling by the TNF receptor superfamily and T cell homeostasis. *Immunity* **13**, 419–422 (2000).
7. Yang, Y., Fang, S., Jensen, J. P., Weissman, A. M. & Ashwell, J. D. Ubiquitin protein ligase activity of IAPs and their degradation in proteasomes in response to apoptotic stimuli. *Science* **288**, 874–877 (2000).
8. Tschopp, J., Martinon, F. & Hofmann, K. Apoptosis: Silencing the death receptors. *Curr. Biol.* **9**, R381–R384 (1999).
9. Van Antwerp, D. J., Martin, S. J., Kafri, T., Green, D. R. & Verma, I. M. Suppression of TNF- α -induced apoptosis by NF- κ B. *Science* **274**, 787–789 (1996).
10. Liu, Z. G., Hsu, H., Goeddel, D. V. & Karin, M. Dissection of TNF receptor 1 effector functions: JNK activation is not linked to apoptosis while NF- κ B activation prevents cell death. *Cell* **87**, 565–576 (1996).
11. Deveraux, Q. L. *et al.* IAPs block apoptotic events induced by caspase-8 and cytochrome *c* by direct inhibition of distinct caspases. *EMBO J.* **17**, 2215–2223 (1998).
12. Roy, N., Deveraux, Q. L., Takahashi, R., Salvesen, G. S. & Reed, J. C. The c-IAP-1 and c-IAP-2 proteins are direct inhibitors of specific caspases. *EMBO J.* **16**, 6914–6925 (1997).
13. Sarin, A., Conan-Cibotti, M. & Henkart, P. A. Cytotoxic effect of TNF and lymphotoxin on T lymphoblasts. *J. Immunol.* **155**, 3716–3718 (1995).
14. Zheng, L. *et al.* Induction of apoptosis in mature T cells by tumour necrosis factor. *Nature* **377**, 348–351 (1995).
15. Duckett, C. S. *et al.* A conserved family of cellular genes related to the baculovirus iap gene and encoding apoptosis inhibitors. *EMBO J.* **15**, 2685–2694 (1996).
16. Chan, F. K. & Lenardo, M. J. A crucial role for p80 TNF-R2 in amplifying p60 TNF-R1 apoptosis signals in T lymphocytes. *Eur. J. Immunol.* **30**, 652–660 (2000).
17. Arch, R. H., Gedrich, R. W. & Thompson, C. B. Translocation of TRAF proteins regulates apoptotic threshold of cells. *Biochem. Biophys. Res. Commun.* **272**, 936–945 (2000).
18. Huang, H. *et al.* The inhibitor of apoptosis, cIAP2, functions as a ubiquitin-protein ligase and promotes *in vitro* ubiquitination of caspases-3 and -7. *J. Biol. Chem.* **275**, 26661–26664 (2000).
19. Rothe, M., Wong, S. C., Henzel, W. J. & Goeddel, D. V. A novel family of putative signal transducers associated with the cytoplasmic domain of the 75 kDa tumor necrosis factor receptor. *Cell* **78**, 681–692 (1994).
20. Rothe, M., Sarma, V., Dixit, V. M. & Goeddel, D. V. TRAF2-mediated activation of NF- κ B by TNF receptor 2 and CD40. *Science* **269**, 1424–1427 (1995).
21. Reinhard, C., Shamo, B., Shyamala, V. & Williams, L. T. Tumor necrosis factor α -induced activation of c-jun N-terminal kinase is mediated by TRAF2. *EMBO J.* **16**, 1080–1092 (1997).
22. Weiss, T. *et al.* Enhancement of TNF receptor p60-mediated cytotoxicity by TNF receptor p80: requirement of the TNF receptor-associated factor-2 binding site. *J. Immunol.* **158**, 2398–2404 (1997).
23. Weiss, T. *et al.* TNFR80-dependent enhancement of TNFR60-induced cell death is mediated by TNFR-associated factor 2 and is specific for TNFR60. *J. Immunol.* **161**, 3136–3142 (1998).
24. Erickson, S. L. *et al.* Decreased sensitivity to tumour-necrosis factor but normal T-cell development in TNF receptor-2-deficient mice. *Nature* **372**, 560–563 (1994).
25. Clem, R. J. *et al.* c-IAP1 is cleaved by caspases to produce a pro-apoptotic C-terminal fragment. *J. Biol. Chem.* **276**, 7602–7608 (2000).
26. Duckett, C. S. & Thompson, C. B. CD30-dependent degradation of TRAF2: implications for negative regulation of TRAF signaling and the control of cell survival. *Genes Dev.* **11**, 2810–2821 (1997).
27. Brown, K. D., Hostager, B. S. & Bishop, G. A. Differential signaling and tumor necrosis factor receptor-associated factor (traf) degradation mediated by CD40 and the Epstein-Barr virus oncoprotein latent membrane protein 1 (LMP1). *J. Exp. Med.* **193**, 943–954 (2001).
28. Peschon, J. J. *et al.* TNF receptor-deficient mice reveal divergent roles for p55 and p75 in several models of inflammation. *J. Immunol.* **160**, 943–952 (1998).
29. Loric, K. L. *et al.* RING fingers mediate ubiquitin conjugating enzyme (E2)-dependent ubiquitination. *Proc. Natl Acad. Sci. USA* **96**, 11364–11369 (1999).
30. Memon, S. A., Petrak, D., Moreno, M. B. & Zacharchuk, C. M. A simple assay for examining the effect of transiently expressed genes on programmed cell death. *J. Immunol. Methods* **180**, 15–24 (1995).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We are grateful to A. Weissman, C. Duckett, and Z. Liu for provision of reagents and reviews of this manuscript. X.L. is a visiting fellow from Bethune International Hospital, China.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.D.A. (e-mail: jda@pop.nci.nih.gov).

correction

Effects of experience and social context on prospective caching strategies by scrub jays

N. J. Emery & N. S. Clayton

Nature **414**, 443–446 (2001).

The exact number of species of American jays is a matter of debate¹. This is particularly evident within the scrub-jay lineage of *Aphelocoma*². Traditionally, *A. coerulescens* included the Florida scrub-jay and the western scrub-jay, among others. It has recently been brought to our attention that the Committee on Classification and Nomenclature of the American Ornithologists' Union now use *A. coerulescens* to refer exclusively to the Florida scrub-jay and *A. californica* to refer to the western scrub-jay². The birds used in the Emery & Clayton³ and Clayton & Dickinson⁴ studies were western scrub-jays: *A. californica*, not *A. coerulescens*. We therefore wish to correct the oversight in this paper and our other papers using these birds, and apologize for any confusion that has arisen. We thank R. Curry, T. Langen and G. Woolfenden for alerting us to the change in nomenclature. □

1. Madge, S. & Burn, H. *Crows and Jays. A Guide to the Crows, Jays and Magpies of the World* (Houghton Mifflin, New York, 1999).
2. American Ornithologists' Union *Check-list of North American Birds* 7th edn (American Ornithologists' Union, Washington DC, 1998).
3. Emery, N. J. & Clayton, N. S. Effects of experience and social context on prospective caching strategies by scrub jays. *Nature* **414**, 443–446 (2001).
4. Clayton, N. S. & Dickinson, A. Episodic-like memory during cache recovery by scrub jays. *Nature* **395**, 272–278 (1998).

erratum

Tyrannosaurus was not a fast runner

John R. Hutchinson & Mariano Garcia

Nature **415**, 1018–1021 (2002).

In this Letter, the row in Table 1 corresponding to 'Gallus, *m_i* (%*m_{body}*)' should read: 1.1, 0.08, 2.0 and 1.5 for the hip, knee, ankle and toe respectively. □

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Chancery Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayourni (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy: Nevin Bayourni (4978)

Scandinavia: Silke Opstrup (4994)

Spain/ Portugal: Leonie Welss
(4954)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel +33 (0) 1 43 20 16 51
Fax +33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel +49 89 54 90 57 11/-2
Fax +49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakata-machi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com
Japan Manager: Kate Cowan

naturejobs

An end to the Irish rover?

Irish scientists have traditionally left their home country to seek research opportunities elsewhere. But two funding schemes are now trying to tempt them back — as well as seeking to attract researchers from other countries, something that Ireland historically has not done on a large scale. The hope is that this initiative will jump-start businesses and diversify the economy. But some fear that the new grants and infrastructure will not be sufficiently tempting — or won't be sustained. The reality might be good job opportunities for scientists willing to take a calculated risk.

The level of investment — a total of about 1.3 billion euros (US\$1.1 billion) over seven years — has certainly caught scientists' attention, both at home and abroad (see pages 4–5). But even Irish scientists who returned home in the mid-1980s and early 1990s, when the country's science funding was particularly grim, view the new-found bounty with guarded optimism. Although they welcome the facilities being built, and hope that the grant schemes will provide the people to staff them, the memory of hard times past serves as a reality check.

Initial results of one scheme, Science Foundation Ireland, bear that caution out. It was set up to attract high-profile scientists — ideally, their entire labs — from other countries. But the results of the first round, awarded last July, are mixed. Of the three biotech-related grants, all went to people already at Irish institutions. Information and communication technology — in which Ireland has a longer history of success — fared a little better. Four of the seven awards went to scientists from abroad — three from England and one from the United States.

But the teething troubles associated with the deployment of new-found finances are problems that many countries would dearly love to have.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

Investing in the future of
the Emerald Isle

p4

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

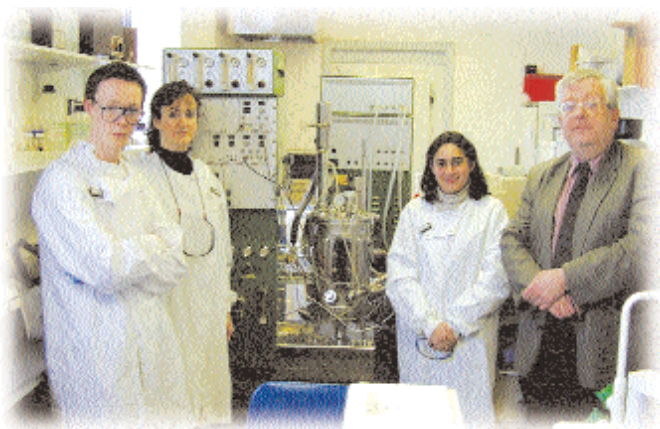
FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Martin Clynes (right) director of the National Institute for Cellular Biotechnology, hopes the newly founded facility will offer long-term career prospects.

Going green Ireland



Reg Shaw: boosting R&D.



After earning his PhD in 1982, Brian Harvey, vice-president of research at University College, Cork, did what many of his fellow Irish scientists have done before and since — he left. Harvey moved to France, mainly because of the limited funding and opportunities at home. He first worked for France's atomic energy commission, then joined its main research agency, the CNRS, eventually becoming a director of research there.

In 1993, he returned — a move that the Irish government hopes more Irish scientists working abroad will undertake. Harvey was head-hunted to run the physiology research centre at University College, Cork. He came home to find the funding scene as bleak as when he left. "Ireland was a desert," he says. "There was no money."

But times have changed. Two government-administered schemes will distribute about 1.2 billion euros (US\$1 billion) by 2007. One, the Programme for Research in Third Level Institutions (PRTL), administered by the Higher Education Authority, was started in 1999 and has so far awarded about 630 million euros, with another 100 million–130 million euros expected during the final round of funding this June. The other scheme, Science Foundation Ireland (SFI), was launched in 2000 by the Department of Enterprise, Trade and Employment, and will give out 635 million euros in grants by 2007.

The government hopes that the schemes will attract researchers of any nationality, including Irish scientists currently working abroad. The programmes will have "a substantial impact on jobs", Harvey says. He anticipates that the number of postdoc slots will double over the next three years under the PRTL funding alone. "We don't produce PhDs fast enough to meet that," he says.

Protein factory

The numbers that managing director Reg Shaw uses to describe his Grange Castle campus, the facility that drug firm Wyeth is building north of Dublin, attract superlatives. The US\$1-billion facility will contain 1,300 employees in 1.2 million square feet.

But in terms of investment in research, the company is initially being conservative. Most of the first wave of employees will be involved in manufacturing protein-

based medicines. After the factory is up and running sometime in 2005, it will start recruiting about 200 scientists for the building's \$300-million research facility.

But because of the huge need for technicians and more skilled scientists, Shaw expects to recruit from both near and far. "Based on the number of expatriates out there, I would think that there would be a number who'd want to come back," he says.

P.S.

Increasing the number of academic staff through the SFI grants presents additional "challenges", Harvey notes, as international candidates need to be sure that the investment will be sustained — especially if they would be moving their entire lab to Ireland.

But Harvey says he is ready to face these challenges. In July, he will move to Dublin to take up the chair of molecular medicine in the newly established Programme for Human Genomics, which, with 44.8 million euros, received the largest single award in the latest round of PRTL grants, announced just before Christmas. The Royal College of Surgeons in Ireland

(RCSI) will lead the project, with Trinity College and University College Dublin (UCD) also participating.

That programme alone will require 40 graduate students, 40 postdocs and 20 academic staff. This growth presents a challenge because Irish science has neither a history of recruiting on that scale, nor one of running such large programmes, says Desmond Fitzgerald, professor of clinical pharmacology at the RCSI. "This has just started," he says. "It's going to take some time to become embedded."

BUILDING UP

John Hegarty, provost of Trinity College, expects new PRTL-funded programmes at Trinity, such as the Centre of Bioengineering (set to receive 5.3 million euros) and the College Institute of Neuroscience (awarded 28 million euros) to help. Such investments are necessary because Ireland lags behind the United States and much of Europe in terms of scientific infrastructure. But buildings alone aren't enough. "We will need more people to fill them up," Hegarty says. And that is where the SFI comes into play.

Ian Robertson, director of Trinity's neuroscience centre, says that the combination of the new building, which is scheduled for completion in three years, and the SFI money has already attracted letters of intent to transfer to Dublin from several prominent neuroscientists in the United States and Europe.

Andy Robertson, director of the UCD's Conway Institute of Biomolecular and Biomedical Research — which was set up using private money and three successive PRTL grants — is equally enthusiastic about building a new centre. The Conway Institute will house bioinformatics, structural biology and proteomics resources, but filling its 300 positions will be a challenge. Attracting people to Ireland is a "marketing and sales exercise that has to go on", Robertson says. The SFI has made some provisions that might help scientists overcome any such reticence — including a programme for one-year stays.

Conway's Robertson, who in 2000 came from the Wellcome Trust in London to start the institute, identifies other possible improvements. One example is help with overheads. If the government would contribute towards the overheads for newly built facilities beyond the first five years, he says, it might convince some international scientists that a move to Ireland need not be short-term.

A SHIFT IN EMPHASIS

Conway's Robertson adds that PRTL-funded centres are supposed to become self-sustaining by attracting external grants — from the Wellcome Trust, for example. But that is easier said than done. There is no single Wellcome centre in Ireland, and only a few Irish researchers get substantial funding from the trust. This could change, says Robertson, but it will require Irish universities to become large-scale research institutions.

Another change that Ireland will need to make is to switch from pharmaceutical manufacturing to research and development (R&D), says Pat Guiry, a chemist at UCD. He notes that nine of the world's largest drug firms have operations in Ireland — but most of them are devoted solely to manufacturing.

That could change if the country can supply the

Housing fits and starts

Dublin's nightlife may make some young researchers consider moving to the city, but its housing prices might give them pause for thought. By some measures, the city is among the 20 most expensive places to live in the world.

One solution for scientists tempted by Ireland's offers but put off by Dublin's prices is to consider Cork or Galway.

University College, Cork, receives the most peer-review research money of any university

in Ireland, and has established four new centres under the latest round of the Programme for Research in Third Level Institutions (PRTL).

The National University of Ireland, Galway, meanwhile, is playing the lead role in a PRTL-funded biomedical engineering programme, which received a further 12.8 million euros (US\$11.3 million) last December. The cost of living in either of these cities can be a fraction of that in Dublin.

P.S.

skills that are in short supply worldwide, such as medicinal chemistry (see *Naturejobs* 4–7; 7 February 2002). There are already signs that a shift to R&D could be successful — drug firm Wyeth is building a US\$1-billion biopharmaceutical campus at Grange Castle near Dublin. Although its main focus will be making protein-based drugs, it will also feature a sizeable R&D laboratory (see 'Protein factory', opposite).

The curriculum at Dublin City University (DCU) seems partly designed to help Ireland move into an R&D environment. Some of its programmes mirror the drug-discovery pipeline, and the campus has several spin-off companies as well as a newly constructed biotech incubator.

The National Institute for Cellular Biotechnology, based at DCU and established with 34.3 million euros in the latest PRTL funding round, should further this trend. Martin Clynes, the institute's director, anticipates that it will include several new faculty members as well as 20–30 students and 20–30 postdocs. He hopes that the new funds will mean better long-term career prospects.

Like Harvey, John Dalton, associate professor of biotechnology at DCU, has seen both sides of progress. When he left a post at the US National Institutes of Health in 1987 in response to the pull of his home country, he returned to laboratories in prefabricated buildings. But in some strange way, that Spartan environment motivated him, ultimately spurring him to attract funds from the World Health Organization to conduct research on malaria.

Irish science is now "seeing lots of money", says Dalton. "In some ways it's a little scary for us." But he and his colleagues hope that the new wealth attracts, rather than intimidates, potential newcomers. ■

Paul Smaglik is editor of *Naturejobs*.

Science Foundation Ireland ♦ www.sfi.ie

Programme for Research in Third Level Institutions

♦ www.heai.ie/projects

Andy Robertson: keen to attract international scientists to Ireland's shores.

